



AFRL-OSR-VA-TR-2013-0012

Information Fusion and Control in Hierarchical Systems

Zhenliang Zhang, Edwin K. P. Chong, Ali Pezeshki, William Moran, and Stephen D. Howard
Colorado State University

May 2013
Final Report

DISTRIBUTION A: Approved for public release.

AIR FORCE RESEARCH LABORATORY
AF OFFICE OF SCIENTIFIC RESEARCH (AFOSR)
ARLINGTON, VIRGINIA 22203
AIR FORCE MATERIEL COMMAND

REPORT DOCUMENTATION PAGE				<i>Form Approved OMB No. 0704-0188</i>	
<small>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</small>					
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.					
1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)

Information Fusion and Control in Hierarchical Systems

Final Report submitted to AFOSR under Contract FA9550-09-1-0518

Zhenliang Zhang, Edwin K. P. Chong, Ali Pezeshki,
William Moran, and Stephen D. Howard

February 27, 2012

Contents

1	Introduction	1
1.1	Background and Related Work	1
2	Balanced Binary Relay Trees	4
2.1	Problem Formulation	4
2.2	Evolution of the Type I and II error probabilities	7
2.3	Error Probability Bounds	9
2.3.1	Case I: analysis as the system propagates from B_m to B_1 . . .	10
2.3.2	Case II: analysis when the system stays inside R	12
2.3.3	Invariant region in B_1	16
2.3.4	Unequal likely hypotheses	18
2.4	Asymptotic Rates	18
2.4.1	Asymptotic decay rate	18
2.4.2	Crummy sensors	20
2.4.3	Fixed point problem	21
2.4.4	Comparison of simulation results	25
3	Sensor Failures and Communication Link Failures	28
3.1	Sensor Failures	28
3.1.1	Problem formulation	28
3.1.2	The evolution of Type I, Type II, and sensor failure error probabilities	30
3.1.3	Error probability bounds	33

3.1.4	Asymptotic rates	38
3.2	Communication Link Failures	39
3.2.1	Problem formulation	39
3.2.2	The evolution of Type I, Type II, and node failure probabilities	41
3.2.3	Asymptotic rates	45
4	Submodularity and Optimality of Fusion Rules	47
4.1	Problem Formulation	47
4.2	Main Results	49
4.2.1	A dynamic programming formulation	49
4.2.2	2-optimal fusion strategy	50
4.2.3	Submodularity	52
5	<i>M</i>-ary Relay Trees	56
5.1	Problem Formulation	56
5.2	Error Probability Bounds	57
5.2.1	Oddary tree	57
5.2.2	Evenary tree	60
5.2.3	Bounds for the total error probability	61
5.2.4	Asymptotic rates	62
5.3	Non-binary Message Alphabets	64
5.4	Scheme Comparison	66
6	Information Geometry of Target Tracking Sensor Networks	69
6.1	Introduction	69
6.2	Target tracking in sensor networks	72
6.3	Principles of Information Geometry	73
6.3.1	Definition of statistical manifold	73
6.3.2	The metric and integrated fisher information distance	74
6.3.3	Geodesics and exponential map	76
6.3.4	Curvatures and information	77

6.4	Application Examples in sensor networks	79
6.4.1	Geodesics and fisher information distance	80
6.4.2	Riemannian and scalar curvatures	88
6.4.3	Ricci curvature tensor field	89
6.5	Statistical Manifold Representation	91
6.5.1	Exponential family of probability density functions	91
6.5.2	Affine immersion of a manifold	93
6.5.3	An example of a gaussian 2-manifold	93
6.5.4	Manifold representations of sensor networks	94
7	Concluding Remarks	98
A	Proof of Proposition 2.3.1	100
B	Proof of Proposition 2.3.2	101
C	Proof of Proposition 2.3.3	102
D	Proof of Proposition 2.3.4	103
E	Proof of Proposition 2.3.5	104
F	Proof of Proposition 2.3.6	106
G	Proof of Proposition 2.3.7	107
H	Proof of Proposition 2.3.8	108
I	Proof of Theorem 2.3.2	109
J	Proof of Proposition 2.3.9	111
K	Proof of Theorem 2.3.3	112
L	Proof of Theorem 2.3.4	114
M	Proof of Proposition 2.3.10	115

Abstract

We consider the distributed detection problem in trees with unbounded height. The first configuration we studied in this report is a balanced binary relay tree, where the leaves of the tree correspond to N identical and independent sensors. Only the leaves are sensors. The root of the tree represents a fusion center that makes the overall detection decision. Each of the other nodes in the tree are relay nodes that combine two binary messages to form a single output binary message. In this way, the information from the sensors is aggregated into the fusion center via the relay nodes. In Chapter II, we assume that the fusion rules are the unit-threshold likelihood-ratio test which are locally optimal in the sense of minimizing the total error probability after fusion. We describe the evolution of the Type I and Type II error probabilities of the binary data as it propagates from the leaves towards the root. Tight upper and lower bounds for the total error probability at the fusion center as functions of N are derived. These characterize how fast the total error probability converges to 0 with respect to N , even if the individual sensors have error probabilities that converge to $1/2$.

In Chapter III, we study the detection performance of balanced binary relay trees where sensors fail with certain probability, in which case we show that the scaling law for the decay rate of the total error probability remains \sqrt{N} . Moreover, we study the case where the communication links in the tree network fail with certain probabilities. Not surprisingly, the step-wise reduction of the total detection error probability is slower than the case where the network has no communication link failures. We show that, under the assumption of identical communication link failure probability in the tree, the exponent of the total error probability at the fusion center is $o(\sqrt{N})$ in the asymptotic regime. In addition, if the given communication link failure probabilities decrease to 0 as communications get closer to the fusion center, then the decay exponent of the total error probability is $\Theta(\sqrt{N})$, provided that the decay of the failure probabilities is sufficiently fast.

In Chapter IV, we call the set of all fusion rules in the tree a fusion strategy. We study the fusion strategy that maximizes the reduction in the total error probability between the sensors and the fusion center. We formulate this optimization problem as a deterministic dynamic program. For trees with finite height, we provide the explicit optimal strategy. Moreover, we show that the reduction in the total error probability is a submodular function. Hence the greedy strategy which only maximizes the level-wise reduction in the total error probability, is close-to the globally optimal strategy in terms of the reduction in the probability of error.

In Chapter V, we consider a more general M -ary relay tree configuration, where each non-leaf node in the tree has M child nodes and only binary messages are allowed to communicate throughout the tree. Similarly we derive tight upper and lower bounds for the Type I and II error probabilities at the fusion center as explicit functions of the number of sensors. These bounds characterize how fast the error probabilities converge to 0 with respect to the number of sensors. Building on the work on the detection performance of M -ary relay trees with binary messages, we further study the case of non-binary relay message alphabets. We characterize the exponent of the error probability with respect to the message alphabet size \mathcal{D} , showing how the detection

performance increases with \mathcal{D} . Our method involves reducing a tree with non-binary relay messages into an equivalent higher-degree tree with only binary messages.

Last, the connections between information geometry and performance of sensor networks for target tracking are explored to pursue a better understanding of placement, planning and scheduling issues. Firstly, the integrated Fisher information distance (IFID) between the states of two targets is analyzed by solving the geodesic equations and is adopted as a measure of target resolvability by the sensor. The differences between the IFID and the well known Kullback-Leibler divergence (KLD) are highlighted. We also explain how the energy functional, which is the “integrated, differential” KLD, relates to the other distance measures. Secondly, the structures of statistical manifolds are elucidated by computing the canonical Levi-Civita affine connection as well as Riemannian and scalar curvatures. We show the relationship between the Ricci curvature tensor field and the amount of information that can be obtained by the network sensors. Finally, an analytical presentation of statistical manifolds as an immersion in the Euclidean space for distributions of exponential type is given. The significance and potential to address system definition and planning issues using information geometry, such as the sensing capability to distinguish closely spaced targets, calculation of the amount of information collected by sensors and the problem of optimal scheduling of network sensor and resources, etc., are demonstrated. The proposed analysis techniques are presented via three basic sensor network scenarios: a simple range-bearing radar, two bearings-only passive sonars, and three ranges-only detectors, respectively.

Chapter 1

Introduction

The *distributed detection* has been studied intensively in the last 30 years. This problem is similar to a classical detection or *hypothesis testing* problem, except that each sensor makes an individual measurement but it is only allowed to communicate a compressed version of its raw measurement. The objective is to choose the most ‘informative’ summarized messages such that a global objective function is optimized. For example, the probability of error for the detection decision is minimized. This problem is also known as *decentralized detection* or *data fusion* in many literatures. This subject is also of interest to the *social learning* perspective, which focus on how fast one peer can learn from other peers in a social network. We begin our discussion with some background and survey of this intriguing problem.

1.1 Background and Related Work

Consider a hypothesis testing problem under two scenarios: *centralized* and *decentralized*. Under the centralized network scenario, all sensors send their raw measurements to the fusion center, which makes a decision based on these measurements. In the decentralized network introduced in [1], sensors send summaries of their measurements and observations to the fusion center. The fusion center then makes a decision. In a decentralized network, information is summarized into smaller messages. Evidently, the decentralized network cannot perform better than the centralized network. It gains because of its limited use of resources and bandwidth; through transmission of summarized information it is more practical and efficient.

The decentralized network in [1] involves the parallel architecture, also known as the *star* architecture [1]–[17], [40], in which all sensors directly connect to the fusion center. Most literatures focus on the issue about how to quantize the measurements so that the probability of error after fusion is minimized. Another perspective is to study how fast the error probability decays with respect to the number of sensors in a large scale network. A typical result is that under the assumption of (conditionally) independence of the sensor observations, the decay rate of the error probability in the parallel architecture is exponential [6].

Several different sensor topologies have been studied under the assumption of conditional independence. The first configuration for such a fusion network considered was the tandem network [18]–[22], [40], in which each non-leaf node combines the information from its own sensor with the message it has received from the node at one level down, which is then transmitted to the node at the next level up. The decay rate of the error probability in this case is sub-exponential [22]. Specifically, as the number of sensors N goes to infinity, the exponent of the error probability is dominated by N^d asymptotically for all $d > 1/2$ [20]. This sensor network represents a situation where the length of the network is the longest possible among all networks with N leaf nodes.

The asymptotic performance of single-rooted tree network with bounded height is discussed in [23]–[31], [40]. Even though the error probability in the parallel configuration decreases exponentially, in a practical implementation, the resources consumed in having each sensor transmit directly to the fusion center might be regarded as excessive. Energy consumption can be reduced by setting up a directed tree, rooted at the fusion center. In this tree structure, measurements are summarized by leaf sensor nodes and sent to their parent nodes, each of which fuses all the messages it receives with its own measurement (if any) and then forwards the new message to its parent node at the next level. This process takes place throughout the tree culminating in the fusion center, where a final decision is made. For bounded-height tree configuration under the Neyman-Pearson criterion, the optimal error exponent is as good as that of the parallel configuration under certain conditions. For example, for a bounded-height tree network with $\lim_{\tau_N \rightarrow \infty} \ell_N / \tau_N = 1$, where τ_N denotes the total number of nodes and ℓ_N denotes the number of leaf nodes, the optimal error exponent is the same as that of the parallel configuration [24], [26]. For bounded-height tree configuration under the Bayesian criterion, the error probability decays exponentially fast to 0 with an error exponent which is worse than the one associated with the parallel configuration [27].

The variation of detection performance with increasing tree height is still largely unexplored. If only the leaf nodes have sensors making observations, and all other nodes simply fuse the messages received and forward the new messages to their parents, the tree network is known as a *relay tree*. The balanced binary relay tree has been addressed in [32], in which it is assumed that the leaf nodes are independent sensors with identical Type I error probability (also known as probability of false alarm, denoted by α_0) and identical Type II error probability (also known as probability of missed detection, denoted by β_0). It is shown there that if the sensor error probabilities satisfy the condition $\alpha_0 + \beta_0 < 1$, then both the Type I and Type II error probabilities at the fusion center converge to 0 as the N goes to infinity. If $\alpha_0 + \beta_0 > 1$, then both the Type I and Type II error probabilities converge to 1, which means that if we flip the decision at the fusion center, then the Type I and Type II error probabilities converge to 0. Because of this symmetry, it suffices to consider the case where $\alpha_0 + \beta_0 < 1$. If $\alpha_0 + \beta_0 = 1$, then the Type I and II error probabilities add up to 1 at each node of the tree. In consequence, this case is not of interest.

In [36], this problem was considered in an M -ary relay tree configuration, where each node with the exception of the sensors has M child nodes. Notice that balanced binary relay trees are simply special cases of M -ary relay trees. To describe the result in [36], let P_N be the total error probability at the fusion center and suppose that each sensor and relay node only transmit binary messages upward to a node at the next level. Then, it is shown in [36] that with any combination of fusion rules, the decay exponent

is upper bounded:

$$\log_2 P_N^{-1} = O(N^{\log_M \frac{(M+1)}{2}}).$$

The case where the relay nodes and the fusion center use the majority dominance rule (with random tie-breaking) to combine messages was also considered in [36], in which case the decay rate of the total error probability is almost optimal. More precisely,

$$\log_2 P_N^{-1} = \Omega(N^{\log_M \lfloor \frac{(M+1)}{2} \rfloor}).$$

Therefore, in the case where M is odd, the majority dominance rule achieves the optimal decay exponent. In the case where M is even, there is a gap between the two bounds. This gap is evident in the case of balanced binary relay tree ($M = 2$), where it is easy to show that the Type I and II error probabilities do not change after fusion with the majority dominance rule. This shows that the lower bound above is tight.

In this report, we develop several new results concerning the detection performance for trees with unbounded height. We show that in trees with unbounded-height, the decay rate of the total error probability is sub-exponentially fast and we show the explicit asymptotic decay exponent.

Chapter 2

Balanced Binary Relay Trees

In this chapter, we consider the balanced binary relay tree configuration and describe the precise evolution of the Type I and Type II error probabilities in this case. In addition, we provide upper and lower bounds for the total error probability at the fusion center as functions of N . These characterize the decay rate of the total error probability. We also show that the total error probability converges to 0 under certain condition even if the sensors are asymptotically crummy, that is, $\alpha_0 + \beta_0 \rightarrow 1$.

2.1 Problem Formulation

We consider the problem of binary hypothesis testing between H_0 and H_1 in a balanced binary relay tree. Leaf nodes are sensors undertaking initial and independent detections of the same event in a scene. These measurements are summarized into binary messages and forwarded to nodes at the next level. Each non-leaf node with the exception of the root, the fusion center, is a relay node, which fuses two binary messages into one new binary message and forwards the new binary message to its parent node. This process takes place at each node culminating in the fusion center, at which the final decision is made based on the information received. Only the leaves are sensors in this tree architecture.

In this configuration, as shown in Fig. 2.1, the closest sensor to the fusion center is as far as it could be, in terms of the number of arcs in the path to the root. In this sense, this configuration is the worst case among all relay trees with N sensors. Moreover, in contrast to the configuration in [24] and [26] discussed earlier, in our balanced binary tree we have $\lim_{\tau_N \rightarrow \infty} \ell_N / \tau_N = 1/2$ (as opposed to 1 in [24] and [26]). Hence, the number of times that information is aggregated is essentially as large as the number of measurements (cf., [24] and [26], in which the number of measurements dominates the number of fusions). In addition, the height of the tree is $\log N$ (log stands for binary logarithm if not specified in this report), which grows as the number of sensors increases.

We assume that all sensors are independent given each hypothesis, and that all sen-

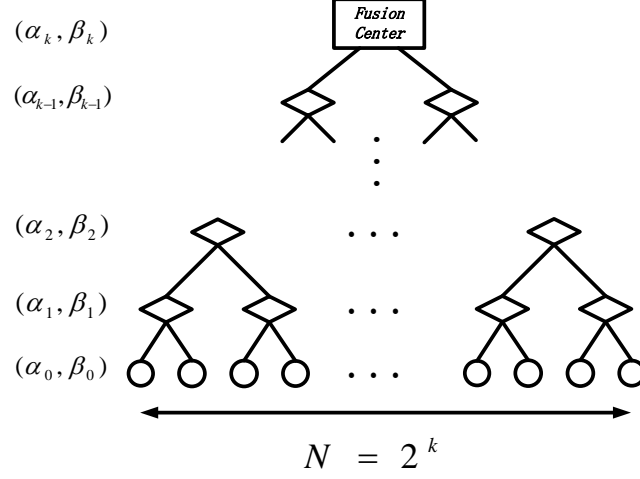


Figure 2.1: A balanced binary relay tree with height k . Circles represent sensors making measurements. Diamonds represent relay nodes which fuse binary messages. The rectangle at the root represents the fusion center making an overall decision.

sors have identical Type I error probability α_0 and identical Type II error probability β_0 . We apply the likelihood-ratio test [41] with threshold 1 as the fusion rule at the relay nodes and at the fusion center. This fusion rule is locally (but not necessarily globally) optimal in the case of equally likely hypotheses H_0 and H_1 ; i.e., it minimizes the total error probability locally at each fusion node. In the case where the hypotheses are not equally likely, the locally optimal fusion rule has a different threshold value, which is the ratio of the two hypothesis probabilities. However, this complicates the analysis without bringing any additional insights. Therefore, for simplicity, we henceforth assume a threshold value of 1 in our analysis. We are interested in following questions:

- What are these Type I and Type II error probabilities as functions of N ?
- Will they converge to 0 at the fusion center?
- If yes, how fast will they converge with respect to N ?

Fusion at a single node receiving information from the two immediate child nodes where these have identical Type I error probabilities α and identical Type II error probabilities β provides a detection with Type I and Type II error probabilities denoted by (α', β') , and given by [32]:

$$(\alpha', \beta') = f(\alpha, \beta) := \begin{cases} (1 - (1 - \alpha)^2, \beta^2), & \alpha \leq \beta, \\ (\alpha^2, 1 - (1 - \beta)^2), & \alpha > \beta. \end{cases} \quad (2.1)$$

Evidently, as all sensors have the same error probability pair (α_0, β_0) , all relay nodes at level 1 will have the same error probability pair $(\alpha_1, \beta_1) = f(\alpha_0, \beta_0)$, and by

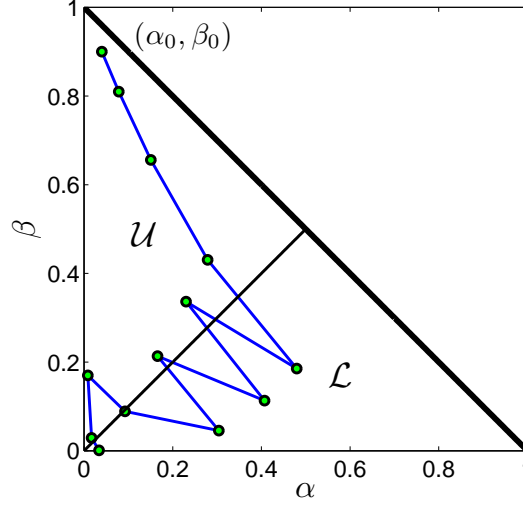


Figure 2.2: A trajectory of the sequence $\{(\alpha_k, \beta_k)\}$ in the (α, β) plane.

recursion,

$$(\alpha_{k+1}, \beta_{k+1}) = f(\alpha_k, \beta_k), \quad k = 0, 1, \dots, \quad (2.2)$$

where (α_k, β_k) is the error probability pair of nodes at the k -th level of the tree.

The recursive relation (2.2) allows us to consider the pair of the Type I and II error probabilities as a discrete dynamic system. In [32], which focuses on the convergence issues for the total error probability, convergence was proved using Lyapunov methods. The analysis of the precise evolution of the sequence $\{(\alpha_k, \beta_k)\}$ and the total error probability decay rate remains open. In this report, we will establish upper and lower bounds for the total error probability and deduce the precise decay rate of the total error probability.

To illustrate the ideas, consider first a single trajectory for the dynamic system given by (2.1), and starting at the initial state (α_0, β_0) . This trajectory is shown in Fig. 2.2. It exhibits different behaviors depending on its distance from the $\beta = \alpha$ line. The trajectory approaches $\beta = \alpha$ very fast initially, but when (α_k, β_k) approaches within a certain neighborhood of the line $\beta = \alpha$, the next pair $(\alpha_{k+1}, \beta_{k+1})$ will appear on the other side of that line. In the next section, we will establish theorems that characterize the precise step-by-step behavior of the dynamic system (2.2). In Section 2.4, we derive upper and lower bounds for (twice) the total error probability P_N at the fusion center as functions of N . These bounds show that the convergence of the total error probability is sub-exponential. Specifically, the exponent of P_N is essentially \sqrt{N} (cf., [24], [26], and [27], where the convergence of the total error probability is exponential in trees with bounded height; more precisely, under the Neyman-Pearson criterion, the optimal error exponent is the same as that of the parallel configuration if leaf nodes dominate; i.e., $\lim_{\tau_N \rightarrow \infty} \ell_N / \tau_N = 1$; but under the Bayesian criterion it is worse).

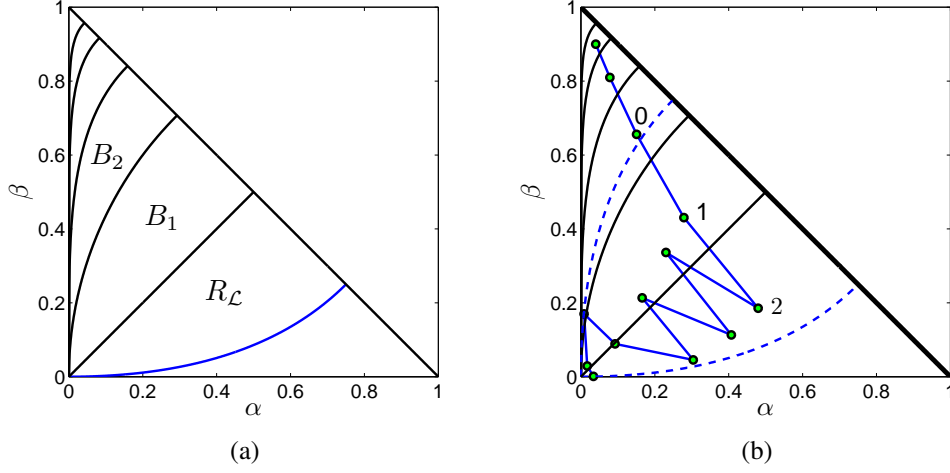


Figure 2.3: (a) Regions B_1 , B_2 , and $R_{\mathcal{L}}$ in the (α, β) plane. (b) The trajectory in Fig. 2.2 superimposed on (a), where solid lines represent boundaries of B_m and dashed lines represent boundaries of R .

2.2 Evolution of the Type I and II error probabilities

The relation (2.1) is symmetric about both of the lines $\alpha + \beta = 1$ and $\beta = \alpha$. Therefore, it suffices to study the evolution of the dynamic system $\{(\alpha_k, \beta_k)\}$ only in the region bounded by $\alpha + \beta < 1$ and $\beta \geq \alpha$. We denote

$$\mathcal{U} := \{(\alpha, \beta) \geq 0 | \alpha + \beta < 1 \text{ and } \beta \geq \alpha\}$$

to be this triangular region. Similarly, define the complementary triangular region

$$\mathcal{L} := \{(\alpha, \beta) \geq 0 | \alpha + \beta < 1 \text{ and } \beta < \alpha\}.$$

We denote the following region by B_1 :

$$B_1 := \{(\alpha, \beta) \in \mathcal{U} | (1 - \alpha)^2 + \beta^2 \leq 1\}.$$

If $(\alpha_k, \beta_k) \in B_1$, then the next pair $(\alpha_{k+1}, \beta_{k+1}) = f(\alpha_k, \beta_k)$ crosses the line $\beta = \alpha$ to the opposite side from (α_k, β_k) . More precisely, if $(\alpha_k, \beta_k) \in \mathcal{U}$, then $(\alpha_k, \beta_k) \in B_1$ if and only if $(\alpha_{k+1}, \beta_{k+1}) = f(\alpha_k, \beta_k) \in \mathcal{L}$. In other words, B_1 is the *inverse image* of \mathcal{L} under mapping f in \mathcal{U} . The set B_1 is shown in Fig. 2.3(a). Fig. 2.3(b) illustrates this behavior of the trajectory for the example in Fig. 2.2. For instance, as shown in Fig. 2.3(b), if the state is at point 1 in B_1 , then it jumps to the next state point 2, on the other side of $\beta = \alpha$.

Denote the following region by B_2 :

$$B_2 := \{(\alpha, \beta) \in \mathcal{U} | (1 - \alpha)^2 + \beta^2 \geq 1 \text{ and } (1 - \alpha)^4 + \beta^4 \leq 1\}.$$

It is easy to show that if $(\alpha_k, \beta_k) \in \mathcal{U}$, then $(\alpha_k, \beta_k) \in B_2$ if and only if $(\alpha_{k+1}, \beta_{k+1}) = f(\alpha_k, \beta_k) \in B_1$. In other words, B_2 is the inverse image of B_1 in \mathcal{U} under mapping f .

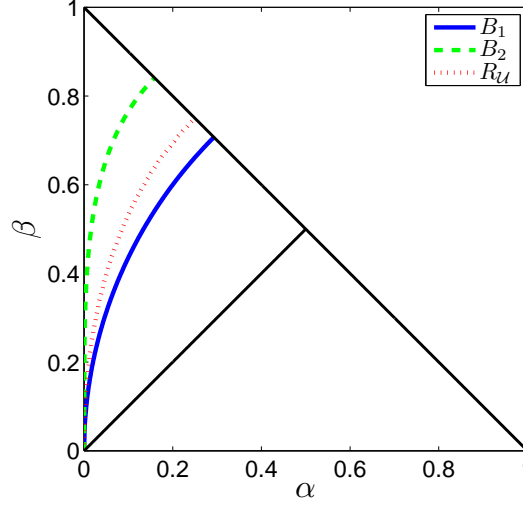


Figure 2.4: Upper boundaries of B_1 , B_2 , and $R_{\mathcal{U}}$.

The behavior of f is illustrated in the movement from point 0 to point 1 in Fig. 2.3(b). The set B_2 is identified in Fig. 2.3(a), lying directly above B_1 .

Now for an integer $m > 1$, recursively define B_m to be the inverse image of B_{m-1} under mapping f , denoted by B_m . It is easy to see that

$$B_m := \{(\alpha, \beta) \in \mathcal{U} \mid (1 - \alpha)^{2^{(m-1)}} + \beta^{2^{(m-1)}} \geq 1 \text{ and } (1 - \alpha)^{2^m} + \beta^{2^m} \leq 1\}.$$

Notice that $\mathcal{U} = \bigcup_{m=1}^{\infty} B_m$. Hence, for any $(\alpha_0, \beta_0) \in \mathcal{U}$, there exists m such that $(\alpha_0, \beta_0) \in B_m$. This gives a complete description of how the dynamics of the system behaves in the upper triangular region \mathcal{U} . For instance, if the initial pair (α_0, β_0) lies in B_m , then the system evolves in the order

$$B_m \rightarrow B_{m-1} \rightarrow \dots \rightarrow B_2 \rightarrow B_1.$$

Therefore, the system will enter B_1 after $m - 1$ levels of fusion; i.e., $(\alpha_{m-1}, \beta_{m-1}) \in B_1$.

As the next stage, we consider the behavior of the system after it enters B_1 . The image of B_1 under mapping f , denoted by $R_{\mathcal{L}}$, is (see Fig. 2.3(a))

$$R_{\mathcal{L}} := \{(\alpha, \beta) \in \mathcal{L} \mid \sqrt{1 - \alpha} + \sqrt{\beta} \geq 1\}.$$

We can define the reflection of B_m about the line $\beta = \alpha$ in the similar way for all m . Similarly, we denote by $R_{\mathcal{U}}$ the reflection of $R_{\mathcal{L}}$ about the line $\beta = \alpha$; i.e.,

$$R_{\mathcal{U}} := \{(\alpha, \beta) \in \mathcal{U} \mid \sqrt{1 - \beta} + \sqrt{\alpha} \geq 1\}.$$

We denote the region $R_{\mathcal{U}} \cup R_{\mathcal{L}}$ by R . We will show that R is an *invariant region* in the sense that once the dynamic system enters R , it stays there. For example, as shown in Fig. 2.3(b), the system after point 1 stays inside R .

Proposition 2.2.1. *If $(\alpha_{k_0}, \beta_{k_0}) \in R$ for some k_0 , then $(\alpha_k, \beta_k) \in R$ for all $k \geq k_0$.*

Proof. First we show that $B_1 \subset R_{\mathcal{U}} \subset B_1 \cup B_2$.

Notice that B_1 , $R_{\mathcal{U}}$, and $B_1 \cup B_2$ share the same lower boundary $\beta = \alpha$. It suffices to show that the upper boundary of $R_{\mathcal{U}}$ lies between the upper boundary of B_2 and that of B_1 (see Fig. 2.4).

First, we show that the upper boundary of $R_{\mathcal{U}}$ lies above the upper boundary of B_1 . We have

$$\begin{aligned} 1 - (1 - \sqrt{\alpha})^2 &\geq \sqrt{1 - (1 - \alpha)^2} \\ \iff 2\sqrt{\alpha} - \alpha &\geq \sqrt{2\alpha - \alpha^2} \\ \iff \alpha^2 + \alpha - 2\alpha^{3/2} &\geq 0, \end{aligned}$$

which holds for all α in $[0, 1)$. Thus, $B_1 \subset R_{\mathcal{U}}$.

Now we prove that the upper boundary of $R_{\mathcal{U}}$ lies below that of B_2 . We have

$$\begin{aligned} (1 - (1 - \alpha)^4)^{1/4} &\geq 1 - (1 - \sqrt{\alpha})^2 \\ \iff 1 - (1 - \alpha)^4 &\geq (2\sqrt{\alpha} - \alpha)^4 \\ \iff -2(\sqrt{\alpha} - 1)^2\alpha(-\alpha^{3/2} + \alpha(\sqrt{\alpha} - 1) \\ &\quad + 4\sqrt{\alpha}(\sqrt{\alpha} - 1) + \alpha - 2) \geq 0, \end{aligned}$$

which holds for all α in $[0, 1)$ as well. Hence, $R_{\mathcal{U}} \subset B_1 \cup B_2$.

Without loss of generality, we assume that $(\alpha_{k_0}, \beta_{k_0}) \in R_{\mathcal{U}}$. It means that $(\alpha_{k_0}, \beta_{k_0}) \in B_1$ or $(\alpha_{k_0}, \beta_{k_0}) \in B_2 \cap R_{\mathcal{U}}$. If $(\alpha_{k_0}, \beta_{k_0}) \in B_1$, then the next pair $(\alpha_{k_0+1}, \beta_{k_0+1})$ lies in $R_{\mathcal{L}}$. If $(\alpha_{k_0}, \beta_{k_0}) \in B_2 \cap R_{\mathcal{U}}$, then $(\alpha_{k_0+1}, \beta_{k_0+1}) \in B_1 \subset R_{\mathcal{U}}$ and $(\alpha_{k_0+2}, \beta_{k_0+2}) \in R_{\mathcal{L}}$. By symmetry considerations, it follows that the system stays inside R for all $k \geq k_0$. □

So far we have studied the precise evolution of the sequence $\{(\alpha_k, \beta_k)\}$ in the (α, β) plane. In the next section, we will consider the step-wise reduction in the total error probability and deduce upper and lower bounds for it.

2.3 Error Probability Bounds

In this section, we will first derive bounds for the total error probability in the case of equally likely hypotheses, where the fusion rule is the likelihood-ratio test with unit threshold. Then we will deduce bounds for the total error probability in the case where the prior probabilities are unequal but the fusion rule remains the same.

The total error probability for a node with (α_k, β_k) is $(\alpha_k + \beta_k)/2$ in the case of equal prior probabilities. Let $L_k = \alpha_k + \beta_k$, namely, twice the total error probability. Analysis of the total error probability results from consideration of the sequence $\{L_k\}$. In fact, we will derive bounds on $\log L_k^{-1}$, whose growth rate is related to the rate of convergence of L_k to 0. We divide our analysis into two parts:

- I We will study the shrinkage of the total error probability as the system propagates from B_m to B_1 ;
- II We will study the shrinkage of the total error probability after the system enters B_1 .

2.3.1 Case I: analysis as the system propagates from B_m to B_1

Suppose that the initial state (α_0, β_0) lies in B_m , where m is a positive integer and $m \neq 1$. From the previous analysis, $(\alpha_{m-1}, \beta_{m-1}) \in B_1$. In this section, we study the rate of reduction of the total error probability as the system propagates from B_m to B_1 .

Proposition 2.3.1. *Suppose that $(\alpha_k, \beta_k) \in B_m$, where m is a positive integer and $m \neq 1$. Then,*

$$1 \leq \frac{L_{k+1}}{L_k^2} \leq 2.$$

The proof is given in Appendix A. Fig. 2.5 shows a plot of values of L_{k+1}/L_k^2 in $\bigcup_{m=2}^{\infty} B_m$. With the recursive relation given in Proposition 2.3.1, we can derive the following bounds for $\log L_k^{-1}$.

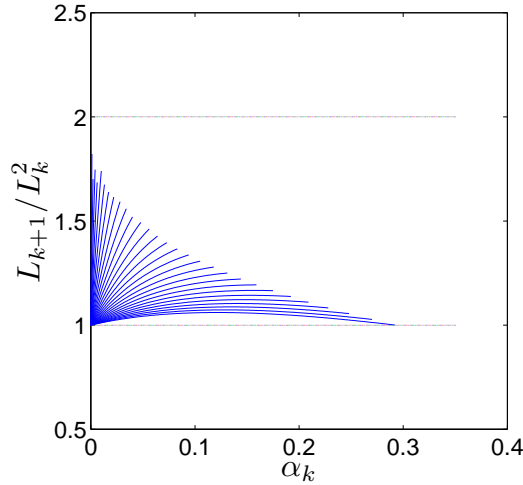


Figure 2.5: Ratio L_{k+1}/L_k^2 in $\bigcup_{m=2}^{\infty} B_m$. Each line depicts the ratio versus α_k for a fixed β_k .

Proposition 2.3.2. *Suppose that $(\alpha_0, \beta_0) \in B_m$, where m is a positive integer and $m \neq 1$. Then, for $k = 1, 2, \dots, m-1$,*

$$2^k (\log L_0^{-1} - 1) \leq \log L_k^{-1} \leq 2^k \log L_0^{-1}.$$

The proof is given in Appendix B. Suppose that the balanced binary relay tree has N leaf nodes. Then, the height of the fusion center is $\log N$. For convenience, let

$P_N = L_{\log N}$ be (twice) the total error probability at the fusion center. Substituting $k = \log N$ into Proposition 2.3.2, we get the following result.

Corollary 2.3.1. *Suppose that $(\alpha_0, \beta_0) \in B_m$, where m is a positive integer and $m \neq 1$. If $\log N < m$, then*

$$N (\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq N \log L_0^{-1}.$$

Notice that the lower bound of $\log P_N^{-1}$ is useful only if $L_0 < 1/2$. Next we derive a lower bound for $\log P_N^{-1}$ which is useful for all $L_0 \in (0, 1)$.

Proposition 2.3.3. *Suppose that $(\alpha_k, \beta_k) \in B_m$, where m is a positive integer and $m \neq 1$. Then,*

$$\frac{L_{k+1}}{L_k^{\sqrt{2}}} \leq 1.$$

The proof is given in Appendix C. Fig. 2.6 shows a plot of values of $L_{k+1}/L_k^{\sqrt{2}}$ in $\bigcup_{m=2}^{\infty} B_m$. With the inequality given in Proposition 2.3.3, we can derive a new lower bound for $\log P_N^{-1}$, which is useful for all $L_0 \in (0, 1)$.

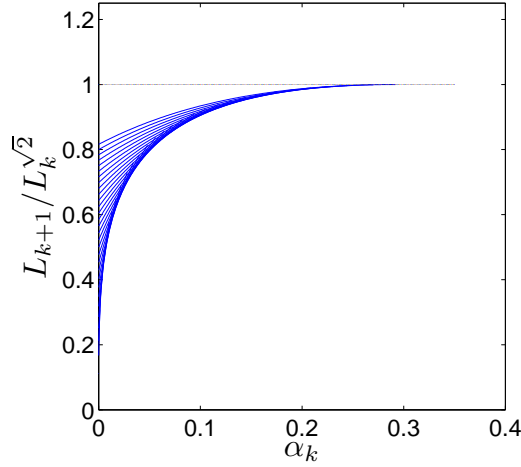


Figure 2.6: Ratio $L_{k+1}/L_k^{\sqrt{2}}$ in $\bigcup_{m=2}^{\infty} B_m$. Each line depicts the ratio versus α_k for a fixed β_k .

Proposition 2.3.4. *Suppose that $(\alpha_0, \beta_0) \in B_m$, where m is a positive integer and $m \neq 1$. If $\log N < m$, then*

$$\log P_N^{-1} \geq \sqrt{N} \log L_0^{-1}.$$

The proof is given in Appendix D.

2.3.2 Case II: analysis when the system stays inside R

We have derived error probability bounds up until the point where the trajectory of the system enters B_1 . In this section, we consider the total error probability reduction from that point on. First we will establish error probability bounds for even-height trees. Then we will deduce error probability bounds for odd-height trees.

Error probability bounds for even-height trees

If $(\alpha_0, \beta_0) \in B_m$ for some $m \neq 1$, then $(\alpha_{m-1}, \beta_{m-1}) \in B_1$. The system afterward stays inside the invariant region R (but not necessarily inside B_1). Hence, the decay rate of the total error probability in the invariant region R determines the asymptotic decay rate. Without loss of generality, we assume that (α_0, β_0) lies in the invariant region R . In contrast to Proposition 2.3.1, which bounds the ratio L_{k+1}/L_k^2 , we will bound the ratio L_{k+2}/L_k^2 associated with taking two steps.

Proposition 2.3.5. *Suppose that $(\alpha_k, \beta_k) \in R$. Then,*

$$1 \leq \frac{L_{k+2}}{L_k^2} \leq 2.$$

The proof is given in Appendix E. Fig. 2.7(a) and Fig. 2.7(b) show plots of values of L_{k+2}/L_k^2 in B_1 and $B_2 \cap R_{\mathcal{U}}$, respectively.

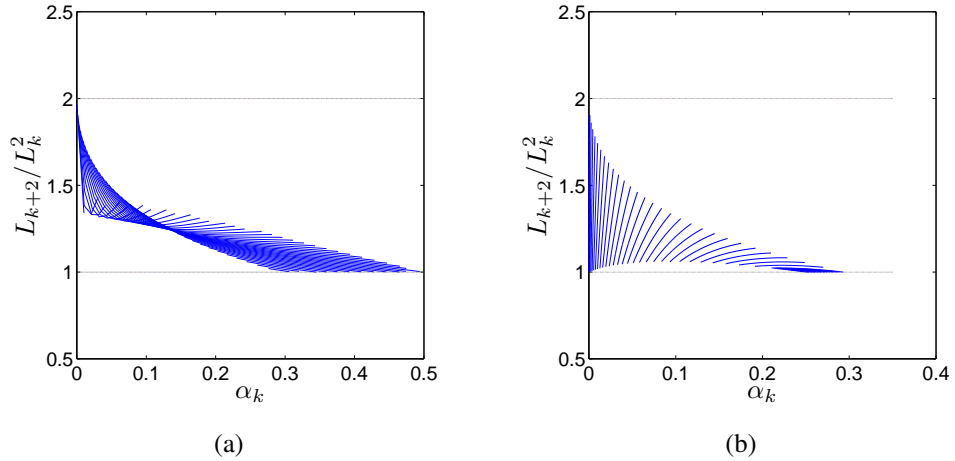


Figure 2.7: (a) Ratio L_{k+2}/L_k^2 in B_1 . (b) Ratio L_{k+2}/L_k^2 in $B_2 \cap R_{\mathcal{U}}$. Each line depicts the ratio versus α_k for a fixed β_k .

Proposition 2.3.5 gives bounds on the relationship between L_k and L_{k+2} in the invariant region R . Hence, in the special case of trees with even height, that is, when $\log N$ is an even integer, it is easy to bound P_N in terms of L_0 . In fact, we will bound $\log P_N^{-1}$ which in turn provides bounds for P_N .

Theorem 2.3.1. *Suppose that $(\alpha_0, \beta_0) \in R$ and $\log N$ is even. Then,*

$$\sqrt{N} (\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq \sqrt{N} \log L_0^{-1}.$$

Proof. If $(\alpha_0, \beta_0) \in R$, then we have $(\alpha_k, \beta_k) \in R$ for $k = 0, 1, \dots, \log N - 2$. From Proposition 2.3.5, we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 2, \dots, \log N - 2$ and some $a_k \in [1, 2]$. Therefore, for $k = 2, 4, \dots, \log N$, we have

$$L_k = a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \cdots a_0^{2^{(k-2)/2}} L_0^{2^{k/2}},$$

where $a_i \in [1, 2]$ for each i . Substituting $k = \log N$, we have

$$\begin{aligned} P_N &= a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \cdots a_0^{2^{(k-2)/2}} L_0^{2^{\log \sqrt{N}}} \\ &= a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \cdots a_0^{\sqrt{N}/2} L_0^{\sqrt{N}}. \end{aligned}$$

Hence,

$$\begin{aligned} \log P_N^{-1} &= -\log a_{(k-2)/2} - 2 \log a_{(k-4)/2} - \cdots \\ &\quad - \frac{\sqrt{N}}{2} \log a_0 + \sqrt{N} \log L_0^{-1}. \end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and $0 \leq \log a_i \leq 1$ for each i . Thus,

$$\log P_N^{-1} \leq \sqrt{N} \log L_0^{-1}.$$

Finally,

$$\begin{aligned} \log P_N^{-1} &\geq -1 - 2 - \cdots - \frac{\sqrt{N}}{2} + \sqrt{N} \log L_0^{-1} \\ &\geq -\sqrt{N} + \sqrt{N} \log L_0^{-1} = \sqrt{N} (\log L_0^{-1} - 1). \end{aligned}$$

□

Notice the lower bound for $\log P_N^{-1}$ in Theorem 2.3.1 is useful only if $L_0 < 1/2$. We further provide a lower bound for $\log P_N^{-1}$ which is useful for all $L_0 \in (0, 1)$.

Proposition 2.3.6. *Suppose that $(\alpha_k, \beta_k) \in R$. Then,*

$$\frac{L_{k+2}}{L_k^{\sqrt{2}}} \leq 1.$$

The proof is given in Appendix F. Fig. 2.8(a) and Fig. 2.8(b) show plots of the ratio inside B_1 and $B_2 \cap R_{\mathcal{U}}$, respectively. Next we derive a new lower bound for $\log P_N^{-1}$.

Proposition 2.3.7. *Suppose that $(\alpha_0, \beta_0) \in R$ and $\log N$ is even. Then,*

$$\log P_N^{-1} \geq \sqrt[4]{N} \log L_0^{-1}.$$

The proof is given in Appendix G.

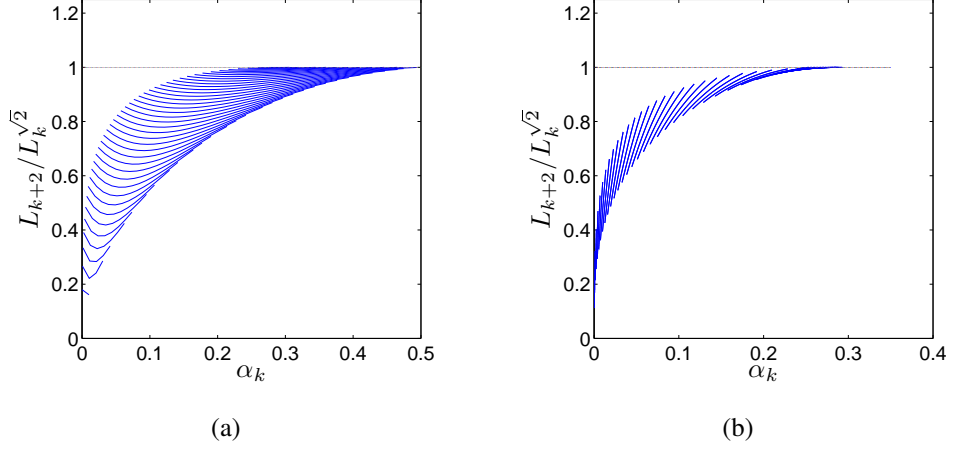


Figure 2.8: (a) Ratio $L_{k+2}/L_k^{\sqrt{2}}$ in B_1 . (b) Ratio $L_{k+2}/L_k^{\sqrt{2}}$ in $B_2 \cap R_{\mathcal{U}}$. Each line depicts the ratio versus α_k for a fixed β_k .

Error probability bounds for odd-height trees

Next we explore the case of trees with odd height; i.e., $\log N$ is an odd integer. Assume that (α_0, β_0) lies in the invariant region R . First, we will establish general bounds for odd-height trees. Then we deduce bounds for the case where there exists $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$ for some $m \in \{0, 1, \dots, \log N - 2\}$.

For odd-height trees, we need to know how much the total error probability is reduced by moving up one level in the tree.

Proposition 2.3.8. *Suppose that $(\alpha_k, \beta_k) \in \mathcal{U}$. Then,*

$$1 \leq \frac{L_{k+1}}{L_k^2}$$

and

$$\frac{L_{k+1}}{L_k} \leq 1.$$

The proof is given in Appendix H. Fig. 2.9(a) and Fig. 2.9(b) show plots of values of L_{k+1}/L_k^2 and L_{k+1}/L_k in \mathcal{U} .

Using Propositions 2.3.5 and 2.3.8, we are about to calculate error probability bounds for odd-height trees as follows.

Theorem 2.3.2. *Suppose that $(\alpha_0, \beta_0) \in R$ and $\log N$ is odd. Then*

$$\sqrt{\frac{N}{2}} (\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq \sqrt{2N} \log L_0^{-1}.$$

The proof is given in Appendix I. Next we consider the special case where there exists $m \in \{0, 1, \dots, \log N - 2\}$ such that $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$.

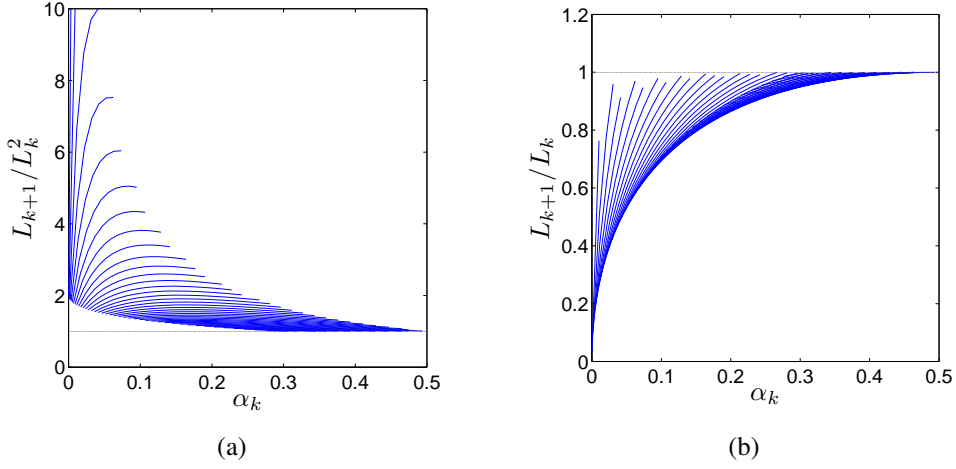


Figure 2.9: (a) Ratio L_{k+1}/L_k^2 in \mathcal{U} . (b) Ratio L_{k+1}/L_k in \mathcal{U} . Each line depicts the ratio versus α_k for a fixed β_k .

Proposition 2.3.9. *Suppose that $(\alpha_k, \beta_k) \in B_1$ and $(\alpha_{k-1}, \beta_{k-1}) \in B_2 \cap R_{\mathcal{U}}$. Then,*

$$\frac{1}{2} \leq \frac{L_{k+1}}{L_k} \leq 1.$$

The proof is given in Appendix J. Fig. 2.10 shows a plot of values of L_{k+1}/L_k in this case.

We have proved in Proposition 2.3.5 that if (α_k, β_k) is in $B_2 \cap R_{\mathcal{U}}$, then the ratio $L_{k+2}/L_k^2 \in [1, 2]$. However, if we analyze each level of fusion, it can be seen that the total error probability decreases exponentially fast from $B_2 \cap R_{\mathcal{U}}$ to B_1 (Proposition 2.3.1). Proposition 2.3.9 tells us that the fusion from B_1 to $R_{\mathcal{C}}$ is a bad step, which does not contribute significantly in decreasing the total error probability.

We can now provide bounds for the total error probability at the fusion center.

Theorem 2.3.3. *Suppose that $(\alpha_0, \beta_0) \in R$, $\log N$ is an odd integer, and there exists $m \in \{0, 1, \dots, \log N - 2\}$ such that $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$.*

If m is even, then

$$\sqrt{2N} (\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq \sqrt{2N} \log L_0^{-1}.$$

If m is odd, then

$$\sqrt{\frac{N}{2}} (\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq \sqrt{\frac{N}{2}} \log L_0^{-1} + \sqrt{\frac{N}{2^{m+2}}}.$$

The proof is given in Appendix K. Finally, by combining all of the analysis above for step-wise reduction of the total error probability, we can write general bounds when the initial error probability pair (α_0, β_0) lies inside B_m , where $m \neq 1$.

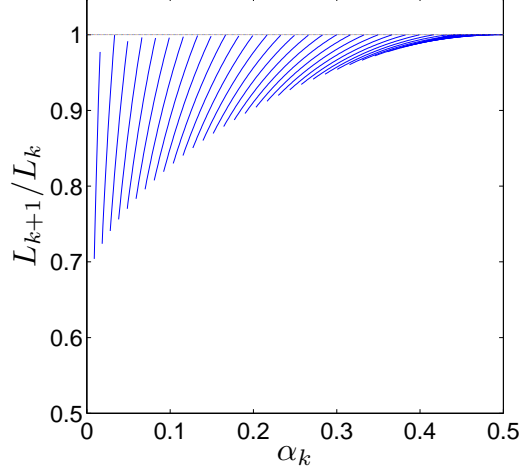


Figure 2.10: Ratio L_{k+1}/L_k in the region $f(B_2 \cup R_U)$. Each line depicts the ratio versus α for a fixed β .

Theorem 2.3.4. Suppose that $(\alpha_0, \beta_0) \in B_m$, where m is an integer and $m \neq 1$.

If $\log N < m$, then (Corollary 2.3.1)

$$N (\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq N \log L_0^{-1}.$$

If $\log N \geq m$, and $\log N - m$ is odd, then

$$\sqrt{2^{m-1}N} (\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq \sqrt{2^{m-1}N} \log L_0^{-1}.$$

If $\log N \geq m$, and $\log N - m$ is even, then

$$\sqrt{2^{m-2}N} (\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq \sqrt{2^{m-2}N} \log L_0^{-1}.$$

The proof uses similar arguments as that of Theorem 2.3.1 and it is provided in Appendix L.

Remark: Notice again that the lower bounds for $\log P_N^{-1}$ above are useful only if $L_0 < 1/2$. However, similar to Proposition 2.3.7, we can derive a lower bound for $\log P_N^{-1}$, which is useful for all $L_0 \in (0, 1)$. It turns out that this lower bound differs from that in Proposition 2.3.7 by a constant term. Therefore, it is omitted.

2.3.3 Invariant region in B_1

Consider the region $\{(\alpha, \beta) \in \mathcal{U} | \beta \leq \sqrt{\alpha} \text{ and } \beta \geq 1 - (1 - \alpha)^2\}$, which is a subset of B_1 (see Fig. 2.11(a)). Denote the union of this region and its reflection with respect to $\beta = \alpha$ by S . It turns out that S is also invariant.

Proposition 2.3.10. *If $(\alpha_{k_0}, \beta_{k_0}) \in S$, then $(\alpha_k, \beta_k) \in S$ for all $k \geq k_0$.*

The proof is given in Appendix M. Fig. 2.11(b) shows a single trajectory of the dynamic system which stays inside S .

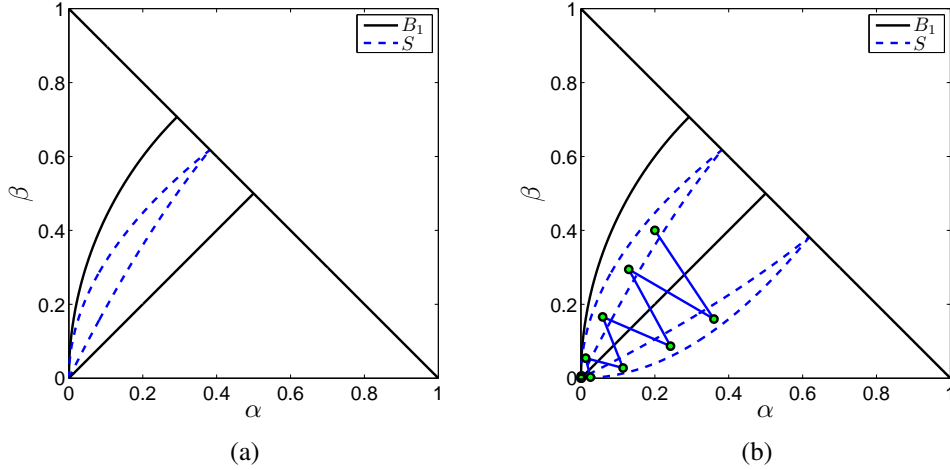


Figure 2.11: (a) Invariant region S (between dashed lines) lies inside B_1 (between solid lines). (b) A trajectory of the system which stays inside S .

We have given bounds for P_N , which is (twice) the total error probability. It turns out that for the case where $(\alpha_0, \beta_0) \in S$, we can bound the Type I and Type II errors individually.

Proposition 2.3.11. *If $(\alpha_k, \beta_k) \in S$, then*

$$1 \leq \frac{\alpha_{k+2}}{\alpha_k^2} \leq 4$$

and

$$1 \leq \frac{\beta_{k+2}}{\beta_k^2} \leq 4.$$

The proof is omitted.

Remark: It is easy to see that as long as the system stays inside B_1 , then in a similar vein, these ratios α_{k+2}/α_k^2 and β_{k+2}/β_k^2 are lower bounded by 1 and upper bounded by a constant. But recall that B_1 is not an invariant region. Thus, it is more interesting to consider S .

Proofs are omitted because they are along similar lines to those in the other proofs. As before, these inequalities give rise to bounds on sequences $\{\alpha_k\}$ and $\{\beta_k\}$. For example, for $\{\alpha_k\}$, we have the following.

Corollary 2.3.2. *If $(\alpha_0, \beta_0) \in S$ and k is even, then*

$$2^{k/2} (\log \alpha_0^{-1} - 2) \leq \log \alpha_k^{-1} \leq 2^{k/2} \log \alpha_0^{-1}.$$

2.3.4 Unequal likely hypotheses

In this section we consider the situation of unequally likely hypotheses; that is, $P(H_0) \neq P(H_1)$. Suppose that the fusion rule is as before: the likelihood ratio test with unit threshold. The resulting total error probability for the nodes at level k is equal to $\hat{L}_k = P(H_0)\alpha_k + P(H_1)\beta_k$, and the total error probability at the fusion center is $\hat{P}_N = \hat{L}_{\log N}$. We are interested in bounds for \hat{P}_N .

Because the fusion rule is the same as before, the previous bounds for $\log L_k^{-1}$ hold. From these bounds, we now derive bounds for \hat{P}_N . Without loss of generality, we assume that $P(H_0) \leq P(H_1)$. We obtain the following:

$$P(H_0)L_k \leq P(H_0)\alpha_k + P(H_1)\beta_k \leq P(H_1)L_k.$$

From these inequalities, we can derive upper and lower bounds for $\log \hat{P}_N^{-1}$. For example, in the case where $(\alpha_0, \beta_0) \in R$ and $\log N$ is even (even-height tree), from Theorem 2.3.1, we have

$$\sqrt{N}(\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq \sqrt{N} \log L_0^{-1},$$

from which we obtain

$$\sqrt{N}(\log L_0^{-1} - 1) + \log P(H_1)^{-1} \leq \log \hat{P}_N^{-1} \leq \sqrt{N} \log L_0^{-1} + \log P(H_0)^{-1}.$$

We have derived error probability bounds for balanced binary relay trees under several scenarios. In the next section, we will use these bounds to study the asymptotic rate of convergence.

2.4 Asymptotic Rates

The asymptotic decay rate of the total error probability with respect to N is considered while the performance of the sensors is constant is the first problem to be tackled. Then we allow the sensors to be asymptotically crummy, in the sense that $\alpha_0 + \beta_0 \rightarrow 1$. We prove that the total error probability still converges to 0 under certain conditions. Last, we will compare the detection performance by applying different strategies in balanced binary relay trees.

In this section, we use the following notation: for positive functions f and g defined on the positive integers, if there exist positive constants c_1 and c_2 such that $c_1 g(N) \leq f(N) \leq c_2 g(N)$ for all sufficiently large N , then we write $f(N) = \Theta(g(N))$. For $N \rightarrow \infty$, the notation $f(N) \sim g(N)$ means that $f(N)/g(N) \rightarrow 1$, $f(N) = \omega(g(N))$ that $f(N)/g(N) \rightarrow \infty$, and $f(N) = o(g(N))$ that $f(N)/g(N) \rightarrow 0$.

2.4.1 Asymptotic decay rate

Notice that as N becomes large, the sequence $\{(\alpha_k, \beta_k)\}$ will eventually move into the invariant region R at some level and stays inside from that point. Therefore, it suffices to consider the decay rate in the invariant region R . Because error probability bounds

for trees with odd height differ from those of the even-height tree by a constant term, without loss of generality, we will only consider trees with even height.

Proposition 2.4.1. *If $L_0 = \alpha_0 + \beta_0$ is fixed, then*

$$\log P_N^{-1} = \Theta(\sqrt{N}).$$

Proof. If $L_0 = \alpha_0 + \beta_0$ is fixed, then by Proposition 2.3.7 we immediately see that $P_N \rightarrow 0$ as $N \rightarrow \infty$ ($\log P_N^{-1} \rightarrow \infty$) and there exists a finite k such that $L_k < 1/2$. To analyze the asymptotic rate, we may assume that $L_0 < 1/2$. In this case, the bounds in Theorem 2.3.1 show that

$$\log P_N^{-1} = \Theta(\sqrt{N}).$$

□

This implies that the convergence of the total error probability is sub-exponential; more precisely, the exponent is essentially \sqrt{N} .

In the special case where $(\alpha_0, \beta_0) \in S$, the Type I and Type II error probabilities decay to 0 with exponent \sqrt{N} individually. Moreover, it is easy to show that the exponent is still \sqrt{N} even if the prior probabilities are unequal.

Given $L_0 \in (0, 1)$ and $\varepsilon \in (0, 1)$, suppose that we wish to determine how many sensors we need to have so that $P_N \leq \varepsilon$. If $L_0 < 1/2$, then the solution is simply to find an N (e.g., the smallest) satisfying the inequality

$$\sqrt{N} (\log L_0^{-1} - 1) \geq -\log \varepsilon.$$

In consequence, we have

$$N \geq ((\log L_0^{-1} - 1) \log \varepsilon)^2.$$

The smallest N grows like $\Theta((\log \varepsilon)^2)$ (cf., [32], in which the smallest N has a larger growth rate). If $L_0 \geq 1/2$, then by Proposition 2.3.7 we can deduce how many levels k are required so that $L_k < 1/2$:

$$\sqrt[4]{N} \log L_0^{-1} > -\log \frac{1}{2} = 1.$$

Therefore, N has to satisfy

$$N > (\log L_0^{-1})^{-4},$$

which implies that

$$k > 4 \log(\log L_0^{-1})^{-1}.$$

Combining with the above analysis for the case where $L_0 < 1/2$, we can then determine the number of sensors required so that $P_N \leq \varepsilon$.

2.4.2 Crummy sensors

In this part we allow the total error probability of each sensor, denoted by $L_0^{(N)}$, to depend on N but still to be constant across sensors.

If $L_0^{(N)}$ is bounded by some constant $L \in (0, 1)$ for all N , then clearly $P_N \rightarrow 0$. It is more interesting to consider $L_0^{(N)} \rightarrow 1$, which means that sensors are asymptotically crummy.

Proposition 2.4.2. *Suppose that $L_0^{(N)} = 1 - \eta_N$ with $\eta_N \rightarrow 0$.*

- (1) *If $\eta_N \geq c_1/\sqrt[4]{N}$, then $P_N \leq e^{-c_1}$.*
- (2) *If $\eta_N = \omega(1/\sqrt[4]{N})$, then $P_N \rightarrow 0$.*
- (3) *If $\eta_N \leq c_2/\sqrt{N}$, then $P_N \geq e^{-c_2}$.*
- (4) *If $\eta_N = o(1/\sqrt{N})$, then $P_N \rightarrow 1$.*

Proof. First we consider part (1). We have

$$\sqrt[4]{N} \log(L_0^{(N)})^{-1} = -\sqrt[4]{N} \log(1 - \eta_N).$$

But as $x \rightarrow 0$, $-\log(1 - x) \sim x/\ln(2)$, from which we obtain

$$\sqrt[4]{N} \log(L_0^{(N)})^{-1} \sim \eta_N \sqrt[4]{N} / \ln(2).$$

From Proposition 2.3.7, it is easy to see that if we have $\eta_N \geq c_1/\sqrt[4]{N}$, then for sufficiently large N we obtain

$$\log P_N^{-1} \geq \sqrt[4]{N} \log(L_0^{(N)})^{-1} \geq c_1 / \ln(2),$$

that is,

$$P_N \leq 2^{-c_1/\ln(2)} = e^{-c_1}.$$

Moreover, if $\eta_N \sqrt[4]{N} \rightarrow \infty$, that is, $\eta_N = \omega(1/\sqrt[4]{N})$, then $P_N \rightarrow 0$. This finishes the proof for part (2).

Next we consider parts (3) and (4). We have

$$\sqrt{N} \log(L_0^{(N)})^{-1} = -\sqrt{N} \log(1 - \eta_N),$$

from which we obtain

$$\sqrt{N} \log(L_0^{(N)})^{-1} \sim \eta_N \sqrt{N} / \ln(2).$$

From Theorem 2.3.1, it is easy to see that if we have $\eta_N \leq c_2/\sqrt{N}$, then for sufficiently large N we obtain

$$\log P_N^{-1} \leq \sqrt{N} \log(L_0^{(N)})^{-1} \leq c_2 / \ln(2),$$

that is,

$$P_N \geq 2^{-c_2/\ln(2)} = e^{-c_2}.$$

Moreover, if $\eta_N \sqrt{N} \rightarrow \infty$, that is, $\eta_N = o(1/\sqrt{N})$, then $P_N \rightarrow 1$.

□

Using part (3) of the above proposition, we derive a necessary condition for $P_N \rightarrow 0$.

Corollary 2.4.1. *Suppose that $L_0^{(N)} = 1 - \eta_N$ with $\eta_N \rightarrow 0$. Then, $P_N \rightarrow 0$ implies that $\eta_N = \omega(1/\sqrt{N})$.*

2.4.3 Fixed point problem

We find the following mathematical problem during our research. Let $\alpha \in [0, 1]$ be given. Define a sequence of numbers γ_k by

$$\gamma_k = \sqrt{1 - \sqrt{1 - \dots \sqrt{1 - (1 - \dots (1 - (1 - \alpha)^2)^2 \dots)^2}}}$$

where the number of nested squares and square roots is k . So, for example, $\gamma_0 = \alpha$, $\gamma_1 = \sqrt{1 - (1 - \alpha)^2}$, and so on.

Define $\theta = (3 - \sqrt{5})/2$, which is simply the golden ratio minus one, or one minus the reciprocal of the golden ratio. In other words, the following expressions are all equal to the golden ratio: $2 - \theta$, $1/(1 - \theta)$, and $1/\sqrt{\theta}$.

Claim: For $\alpha \in [0, \theta]$,

- the odd subsequence of $\{\gamma_k\}$ converges from above to $\sqrt{\alpha}$, and
- the even subsequence of $\{\gamma_k\}$ converges from below to $1 - (1 - \alpha)^2$.

For $\alpha \in [\theta, 1]$,

- the odd subsequence of $\{\gamma_k\}$ converges from above to $1 - (1 - \alpha)^2$, and
- the even subsequence of $\{\gamma_k\}$ converges from below to $\sqrt{\alpha}$.

Equivalently, for each $\alpha \in [0, 1]$, the odd subsequence converges from above to $\max(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$, and the even subsequence converges from below to $\min(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$.

We have no direct proof of this claims. However, what we can easily show is that the odd subsequence is monotone decreasing and bounded below by $\max(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$, and that the even subsequence is monotone increasing and bounded above by $\min(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$.

Functional Representation of Problem

Define the function $r_k : [0, 1] \rightarrow [0, 1]$ such that $\gamma_k = r_k(\alpha)$. Then, it is apparent that r_k satisfies the recursion

$$r_{k+1}(\alpha) = \sqrt{1 - r_k((1 - \alpha)^2)}, \quad \alpha \in [0, 1], \quad (2.3)$$

with $r_0(\alpha) = \alpha$. We can similarly write a recursion relating r_{k+2} to r_k :

$$r_{k+2}(\alpha) = \sqrt{1 - \sqrt{1 - r_k((1 - (1 - \alpha)^2)^2)}}, \quad \alpha \in [0, 1]. \quad (2.4)$$

Starting with $r_0(\alpha) = \alpha$ gives rise to the even subsequence, and starting with $r_1(\alpha) = \sqrt{1 - (1 - \alpha)^2}$ gives rise to the odd subsequence. The convergence claim before amounts to pointwise convergence claims about these subsequences of functions. But since we have a recursion for them, they can only converge to fixed points of the recursion (2.4).

More specifically, $r : [0, 1] \rightarrow [0, 1]$ is a *fixed point* of the first recursion (2.3) if and only if

$$r(\alpha) = \sqrt{1 - r((1 - \alpha)^2)}, \quad \alpha \in [0, 1]. \quad (2.5)$$

With some algebraic manipulations, we can rewrite this as

$$r(\alpha) = 1 - r(1 - \sqrt{\alpha})^2, \quad \alpha \in [0, 1]. \quad (2.6)$$

Similarly, for a fixed point of (2.4),

$$r(\alpha) = \sqrt{1 - \sqrt{1 - r((1 - (1 - \alpha)^2)^2)}}, \quad \alpha \in [0, 1]. \quad (2.7)$$

If we can show that the two desired functions $\max(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$ and $\min(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$ are the only “legitimate” solutions of (2.7), then we are done.

Write the recursion for $\{r_k\}$ as $r_{k+1} = \Phi r_k$, where Φ is the operator defined by (2.3). The recursion for the even and odd subsequences is $r_{k+2} = \Phi^2 r_k$, where Φ^2 means Φ composed with itself. When we speak of a *fixed point* of Φ , we mean a function $r : [0, 1] \rightarrow [0, 1]$ such that $r = \Phi r$. Two functions ρ_1 and ρ_2 constitute an *orbit of period 2* of Φ if $\rho_1 = \Phi \rho_2$ and $\rho_2 = \Phi \rho_1$.

Some observations:

- Any fixed point of Φ is also a fixed point of Φ^2 . Moreover, two functions that form an orbit of Φ of period 2 are both fixed points of Φ^2 .
- Conversely, any fixed point of Φ^2 is either a fixed point or a point on an orbit of Φ of period 2. For example, the constant functions 0 and 1 are fixed points of Φ^2 , but they are not fixed points of Φ . Instead, they constitute an orbit of period 2.
- The functions on $[0, 1]$ given by $\sqrt{\alpha}$ and $1 - (1 - \alpha)^2$ are both fixed points of Φ (and hence also Φ^2). Moreover, the constant function $\sqrt{\theta}$ (which is also equal to $1 - \theta$, $1 - (1 - \theta)^2$, and $(1/\theta) - 2$) is a fixed point of Φ (and hence also Φ^2).
- Fixing $\alpha = 0$ in (2.7), we see that $r(0)$ can only take values 0, 1, or $\sqrt{\theta}$. The same can be said for $r(1)$ and $r(\theta)$. Any fixed point must satisfy these constraints at $\alpha = 0, 1$, and θ .
- Φ maps a function on $[0, \theta]$ to a function on $[\theta, 1]$, and vice versa. So if ρ_1 and ρ_2 are fixed points of Φ , then the two functions given by $\rho_1(\alpha)1_{[0, \theta]}(\alpha) + \rho_2(\alpha)1_{(\theta, 1]}(\alpha)$ and $\rho_2(\alpha)1_{[0, \theta]}(\alpha) + \rho_1(\alpha)1_{(\theta, 1]}(\alpha)$ constitute an orbit of Φ of period 2. These are functions obtained from ρ_1 and ρ_2 by “swapping” them on the interval $[0, \theta]$.

- Similarly, if ρ_1 and ρ_2 constitute an orbit of Φ , then the two functions given by $\rho_1(\alpha)1_{[0,\theta]}(\alpha) + \rho_2(\alpha)1_{(\theta,1]}(\alpha)$ and $\rho_2(\alpha)1_{[0,\theta]}(\alpha) + \rho_1(\alpha)1_{(\theta,1]}(\alpha)$ are both fixed points of Φ . For example, both functions $1_{[0,\theta]}$ and $1_{(\theta,1]}$ are fixed points of Φ .
- Φ^2 maps a function on $[0, \theta]$ to a function on $[0, \theta]$ (and similarly on $[\theta, 1]$). Hence, for Φ^2 , we can consider the intervals $[0, \theta]$ and $[\theta, 1]$ separately and independently. In particular, if ρ_1 and ρ_2 are fixed points of Φ^2 , then so are the functions given by $\rho_1(\alpha)1_{[0,\theta]}(\alpha) + \rho_2(\alpha)1_{(\theta,1]}(\alpha)$ and $\rho_2(\alpha)1_{[0,\theta]}(\alpha) + \rho_1(\alpha)1_{(\theta,1]}(\alpha)$.
- Finally, given two functions ρ_1 and ρ_2 that either are both fixed points or constitute an orbit of period 2 of Φ , we can construct two fixed points of Φ^2 : $\rho_1(\alpha)1_{[0,\theta]}(\alpha) + \rho_2(\alpha)1_{(\theta,1]}(\alpha)$ and $\rho_2(\alpha)1_{[0,\theta]}(\alpha) + \rho_1(\alpha)1_{(\theta,1]}(\alpha)$.
The two fixed points of Φ^2 of interest in our original claim are $\sqrt{\alpha}1_{[0,\theta]}(\alpha) + (1 - (1 - \alpha)^2)1_{(\theta,1]}(\alpha)$ and $(1 - (1 - \alpha)^2)1_{(\theta,1]}(\alpha) + \sqrt{\alpha}1_{[0,\theta]}(\alpha)$, which are $\max(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$ and $\min(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$, respectively.

Other Fixed Points

Define the function $F : [0, 1] \rightarrow [0, 1]$ by $F(x) = 1 - x$ (the “flip” function), and the function $S : [0, 1] \rightarrow [0, 1]$ by $S(x) = x^2$ (the “square” function). Then, for $r : [0, 1] \rightarrow [0, 1]$, the function Φr can be written as $S^{-1}FrSF$ (in operator notation, so that SF , say, means S composed with F). Hence, r is a fixed point of Φ if and only if

$$r = S^{-1}FrSF, \quad (2.8)$$

which corresponds to (2.5). Algebraic manipulations are much easier using this notation. For example, note that $F^{-1} = F$, from which we can easily derive

$$r = FSrFS^{-1}, \quad (2.9)$$

which corresponds to (2.6). The two-step version (2.7) is also easy to write:

$$r = (S^{-1}F)^2r(SF)^2, \quad (2.10)$$

Not that the two functions $\sqrt{\alpha}$ and $1 - (1 - \alpha)^2$ in this new notation are S^{-1} and FSF , respectively.

To simplify the calculations, substitute $\sigma = rS$ in (2.8). Then,

$$\sigma(S^{-1}F) = (S^{-1}F)\sigma.$$

In other words, $r = \sigma S^{-1}$ satisfies (2.8) if and only if σ commutes with $(S^{-1}F)$. Note that $S^{-1}F(\sqrt{\theta}) = S^{-1}F(1 - \theta) = \sqrt{\theta}$, which shows that the constant function $\sqrt{\theta}$ commutes with $(S^{-1}F)$. It is clear that any power of $(S^{-1}F)$ commutes with itself. From this, we generate an infinite family of such fixed points r :

$$r = (S^{-1}F)^n S^{-1}, \quad n \in \mathbb{Z}. \quad (2.11)$$

The two special functions S^{-1} and FSF identified before are special cases with $n = 0$ and $n = -2$, respectively.

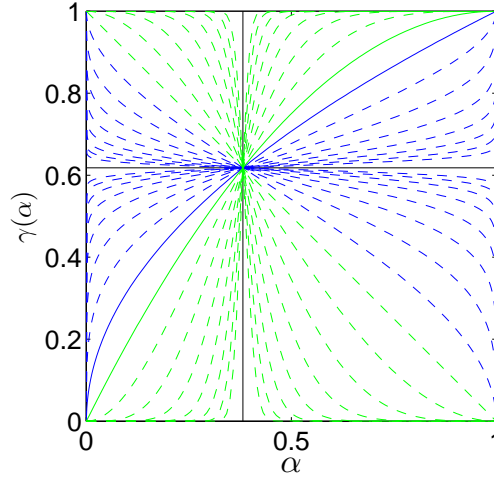


Figure 2.12: Functions in (2.11).

We could similarly make the substitution $\tau = Fr$ in (2.8), leading to $\tau(SF) = (SF)\tau$. This gives rise to the family

$$r = F(SF)^m, \quad m \in \mathbb{Z}.$$

But note that this family reduces to the previous one by substituting $m = -n - 1$. Similarly, working with (2.9) does not generate any new fixed points.

If we consider the two-step version, we arrive at the conclusion that $r = \sigma S^{-1}$ satisfies (2.10) if and only if σ commutes with $(S^{-1}F)^2$. Naturally, the functions in (2.11) and the constant function $\sqrt{\theta}$ also satisfy (2.10). Moreover, because $0 = (S^{-1}F)^2(0)$ and $1 = (S^{-1}F)^2(1)$, the constant functions 0 and 1 also clearly satisfy (2.10).

Figure 2.12 shows plots of several functions of interest on $[0, 1]$. The blue and green plots are the functions in (2.11) for nonnegative and negative n , respectively. We have pointed out the two cases $n = 0$ (i.e., $\sqrt{\alpha}$) and $n = -2$ (i.e., $1 - (1 - \alpha)^2$) using solid lines (in contrast to the other dashed lines). The solid black vertical line is at $\alpha = \theta$. The solid black horizontal line is the constant function $\sqrt{\theta}$. Notice that all the blue and green plots intersect at $(\theta, \sqrt{\theta})$. The plots in Figure 2.12 depict a large family of fixed points of (2.7), recalling that we can combine blue and green plots on the intervals $[0, \theta]$ and $(\theta, 1]$.

Figure 2.13 overlays on Figure 2.12 four red plots, depicting the first four iterations $k = 1, \dots, 4$ of (2.3). The two red plots above $\sqrt{\alpha}$ in the region $[0, \theta]$ are the odd iterates ($k = 1, 3$) and the two red plots below $1 - (1 - \alpha)^2$ in the region $[0, \theta]$ are the even iterates ($k = 2, 4$). Note that the solid blue/green plots are the only ones that lie between the red plots. Recall that odd subsequence of (2.7) is monotone decreasing and bounded below by $\max(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$, and that the even subsequence is monotone increasing and bounded above by $\min(\sqrt{\alpha}, 1 - (1 - \alpha)^2)$. Hence, among all the fixed points in Figure 2.12, only these two are legitimate fixed points of (2.7).

Are there other fixed points of (2.7)? Recall that any fixed point $r = \sigma S^{-1}$ is

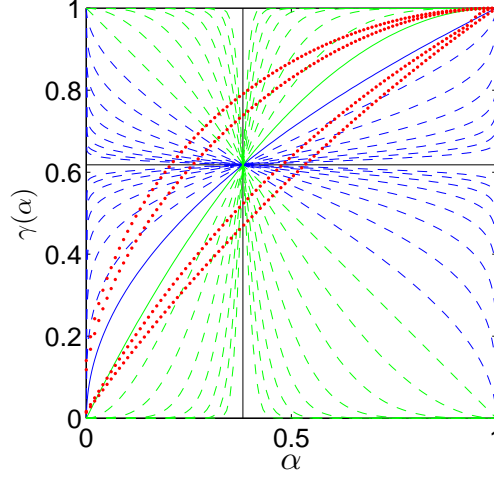


Figure 2.13: Functions in (2.11) together with subsequences of (2.3) in red.

such that σ commutes with $(S^{-1}F)^2$. We have pointed out that any integer power of $(S^{-1}F)$ will do the job.

Are there other functions that commute with $(S^{-1}F)^2$? One candidate is this: Suppose that $(S^{-1}F)^2$ has a *commutable factorization*, which means that $(S^{-1}F)^2 = AB = BA$ for functions $A : [0, 1] \rightarrow [0, 1]$ and $B : [0, 1] \rightarrow [0, 1]$. Then, any integer power of A and B commutes with $(S^{-1}F)^2$. Clearly $A = B = S^{-1}F$ is such an example. But are there others?

2.4.4 Comparison of simulation results

We end this section by comparing the quantitative behavior of the unit-threshold likelihood-ratio rule with that of other fusion rules of interest. First, we define two particular fusion rules that can be applied at an individual node:

- OR rule: the parent node decides 0 if and only if both the child nodes send 0;
- AND rule: the parent node decides 1 if and only if both the child nodes send 1.

Notice that the unit-threshold likelihood-ratio rule reduces to either the AND rule or the OR rule, depending on the values of the Type I and Type II error probabilities at the particular level of the tree. For our quantitative comparison, we consider three system-wide fusion strategies that we will compare with the case that uses the unit-threshold likelihood-ratio rule at every node:

- OR strategy: Every fusion uses the OR rule;
- AND strategy: Every fusion uses the AND rule;

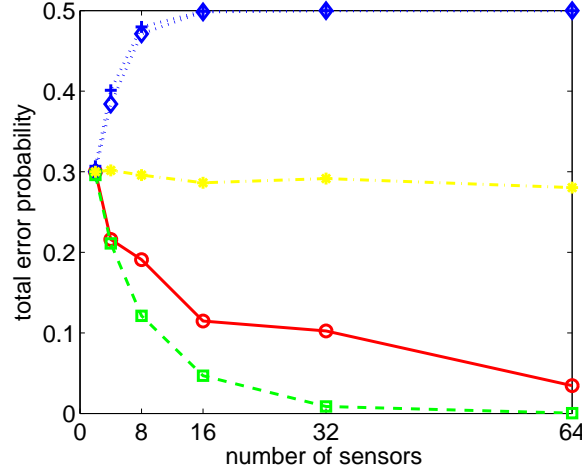


Figure 2.14: Total error probability plots. Dashed line: centralized parallel fusion strategy. Solid line: unit-threshold likelihood-ratio rule for balanced binary relay tree. Dotted line with ‘ \diamond ’ marker: OR strategy. Dotted line with ‘+’ marker: AND strategy. Dash-dot line: RAND strategy.

- RAND strategy: At each level of the tree, we randomly pick either the AND rule or the OR rule with equal probability, and independently over levels, and apply that rule to all the nodes at that level.

In Fig. 2.14, we show plots of the total error probability as a function of N for the tree that uses the unit-threshold likelihood-ratio rule at every node (the one analyzed in this report). We also plot the total error probabilities for the AND and OR strategies, as well as the average total error probability over 100 independent trials of the RAND strategy. For comparison purposes, we also plot the error probability curve of the centralized parallel fusion strategy.

We can see from Fig. 2.14 that the total error probability for the centralized parallel strategy decays to 0 faster than that of the binary relay tree that uses the unit-threshold likelihood-ratio rule at every node. This is not surprising, because the former is known to be exponential, as discussed earlier, while the latter is sub-exponential with exponent \sqrt{N} , as shown in this report. The AND and OR strategies both result in total error probabilities converging monotonically to $1/2$, while the RAND strategy results in an average total error probability that does not decrease much with N .

We have studied the detection performance of balanced binary relay trees. We precisely describe the evolution of error probabilities in the (α, β) plane as we move up the tree. This allows us to deduce error probability bounds at the fusion center as functions of N under several different scenarios. These bounds show that the total error probability converges to 0 sub-exponentially, with an exponent that is essentially \sqrt{N} . In addition, we allow all sensors to be asymptotically crummy, in which case we deduce the necessary and sufficient conditions for the total error probability to converge to 0. All our results apply not only to the fusion center, but also to any other node in the tree network. In other words, we can similarly analyze a sub-tree inside the original

tree network.

Chapter 3

Sensor Failures and Communication Link Failures

In this Chapter, we further study the detection performance of balanced binary relay trees with sensor failures and communication link failures build on the formulations in Chapter II.

3.1 Sensor Failures

3.1.1 Problem formulation

We keep all the notations and definitions defined in Chapter II. Moreover for the sensor failure case, we assume that all sensors have identical failure probability q_0 . Assuming equal prior probabilities, we use the likelihood-ratio test [41] when fusing binary messages at intermediate relay nodes and the fusion center. Consider the simple problem of fusing binary messages passed to a node by its two immediate child nodes. Assume that the two child nodes have identical Type I error probability α , identical Type II error probability β , and identical failure probability q .

Denote the Type I error, Type II error, and failure probabilities after the fusion by (α', β', q') . This parent node fails to provide any message to the node at the next level if and only if both its child nodes fail to forward any message. Hence, we have

$$q' = q^2. \quad (3.1)$$

If one of the child nodes fails and the other one sends its message to the parent node, then Type I and Type II error probabilities do not change since the parent node receives only one binary message. The probability of this event is $2q(1 - q)$, in which case we have

$$(\alpha', \beta') = (\alpha, \beta). \quad (3.2)$$

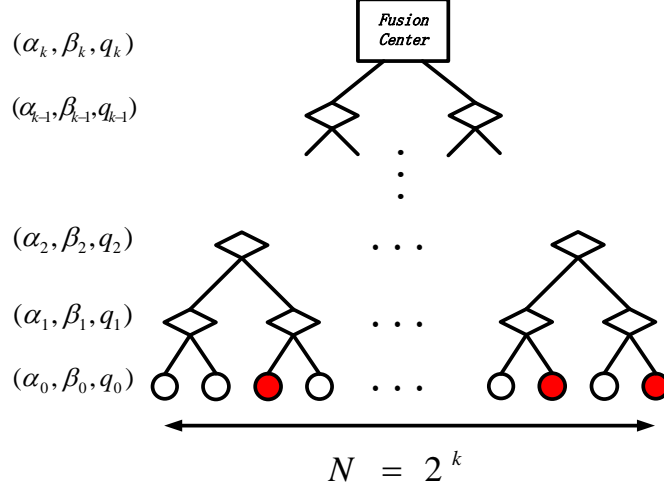


Figure 3.1: A balanced binary relay tree with height k . Circles represent sensors making measurements. Diamonds represent relay nodes which fuse binary messages. The rectangle at the root represents the fusion center making an overall decision.

If both child nodes send their messages to the parent node, then the scenario is the same as that in [32] and [33]. The probability of this event is $(1 - q)^2$, in which case we have

$$(\alpha', \beta') = \begin{cases} (1 - (1 - \alpha)^2, \beta^2), & \alpha \leq \beta, \\ (\alpha^2, 1 - (1 - \beta)^2), & \alpha > \beta. \end{cases} \quad (3.3)$$

Let $\bar{\alpha}'$ and $\bar{\beta}'$ be the mean Type I and Type II error probabilities conditioned on the event that at least one of these child nodes forwards its message to the parent node, i.e., the parent node has data. We have

$$(\bar{\alpha}', \bar{\beta}', q') = f(\alpha, \beta, q) \quad (3.4)$$

$$= \begin{cases} \left(\frac{(1-q)(2\alpha - \alpha^2) + 2q\alpha}{1+q}, \frac{(1-q)\beta^2 + 2q\beta}{1+q}, q^2 \right), & \alpha \leq \beta, \\ \left(\frac{(1-q)\alpha^2 + 2q\alpha}{1+q}, \frac{(1-q)(2\beta - \beta^2) + 2q\beta}{1+q}, q^2 \right), & \alpha > \beta. \end{cases} \quad (3.5)$$

Our assumption is that all sensors have the same error probabilities (α_0, β_0, q_0) . Therefore by (3.5), all relay nodes at level 1 will have the same *error probability triplet* $(\alpha_1, \beta_1, q_1) = f(\alpha_0, \beta_0, q_0)$ (where α_1 and β_1 are the conditional mean error probabilities). Similarly by (3.4), we can calculate error probability triplets for nodes at all other levels. We have

$$(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) = f(\alpha_k, \beta_k, q_k), \quad k = 1, 2, \dots, \quad (3.6)$$

where (α_k, β_k, q_k) is the error probability triplet of nodes at the k th level of the tree. Notice that if we let $q_0 = 0$, then the recursive relation reduces to the recursion in [33].

The relation (3.6) allows us to consider (α_k, β_k, q_k) as a discrete dynamic system. For the case where $q_0 = 0$, we have studied (See [33]) the precise evolution of the sequence $\{(\alpha_k, \beta_k)\}$, derived total error probability bounds as functions of N , and established asymptotic decay rates. In this report, we will study the case where $q_0 \neq 0$. We will derive total error probability bounds and determine the decay rate of the total error probability.

To develop intuition, let us start by looking at the single trajectory shown in Fig. 3.2(a), starting at the initial state (α_0, β_0, q_0) . We observe that q_k decreases very fast to 0. In addition, as shown in Fig. 3.2(b), the trajectory approaches $\beta = \alpha$ at the beginning. After (α_k, β_k) gets too close to $\beta = \alpha$, the next pair $(\alpha_{k+1}, \beta_{k+1})$ will be repelled toward the other side of the line $\beta = \alpha$. This behavior is similar to the scenario where $q = 0$. For the case where $q = 0$, there exist an invariant region in the sense that the system stays in the invariant region once the system enters it [33]. Is there an invariant region for the case where $q \neq 0$? We answer this question by precisely describing this invariant region in \mathbb{R}^3 .

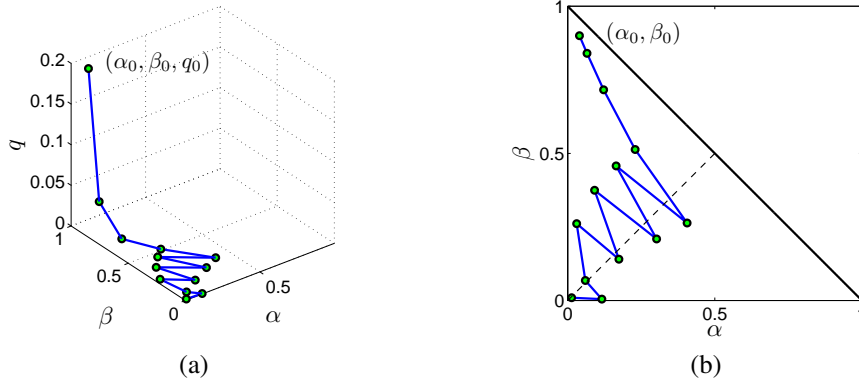


Figure 3.2: (a) A typical trajectory of (α_k, β_k, q_k) in the (α, β, q) coordinates. (b) The trajectory in (a) projected onto the (α, β) plane.

3.1.2 The evolution of Type I, Type II, and sensor failure error probabilities

The relation (3.5) is symmetric about the hyper-planes $\alpha + \beta = 1$ and $\beta = \alpha$. Thus, it suffices to study the evolution of the dynamic system only in the region bounded by $\alpha + \beta < 1$, $\beta \geq \alpha$, and $0 \leq q \leq 1$. Let $\mathcal{U} := \{(\alpha, \beta, q) \geq 0 | \alpha + \beta < 1, \beta \geq \alpha, \text{ and } 0 \leq q \leq 1\}$ be this triangular prism. Similarly, define the complementary triangular prism $\mathcal{L} := \{(\alpha, \beta, q) \geq 0 | \alpha + \beta < 1, \beta < \alpha, \text{ and } 0 \leq q \leq 1\}$.

First, we denote the following region by

$$B_1 := \{(\alpha, \beta, q) \in \mathcal{U} | \beta \leq (-q + \sqrt{q^2 + (1-q)^2(2\alpha - \alpha^2) + 2q(1-q)\alpha}) / (1-q)\}.$$

If $(\alpha_k, \beta_k, q_k) \in B_1$, then the next pair $(\alpha_{k+1}, \beta_{k+1}, q_{k+1})$ jumps across the plane $\beta =$

α away from (α_k, β_k, q_k) . More precisely, if $(\alpha_k, \beta_k, q_k) \in \mathcal{U}$, then $(\alpha_k, \beta_k, q_k) \in B_1$ if and only if $(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) \in \mathcal{L}$.

It is easy to see from (3.5) and (3.6) that, if we start with $(\alpha_0, \beta_0, q_0) \in \mathcal{U} \setminus B_1$, then before the system enters B_1 , we have $\alpha_{k+1} > \alpha_k$ and $\beta_{k+1} < \beta_k$. Thus, the system moves towards the $\beta = \alpha$ plane. Therefore, if the number of sensors N is sufficiently large, then the system is guaranteed to enter B_1 .

Next we consider the behavior of the system after it enters B_1 . If $(\alpha_k, \beta_k, q_k) \in B_1$, we consider the position of the next pair $(\alpha_{k+1}, \beta_{k+1}, q_{k+1})$, i.e., consider the *image* of B_1 under f , denoted by $R_{\mathcal{L}}$. Similarly we denote the reflection of $R_{\mathcal{L}}$ with respect to $\beta = \alpha$ by $R_{\mathcal{U}}$. We find that

$$R_{\mathcal{U}} := \{(\alpha, \beta, q) \in \mathcal{U} \mid \beta \leq -\alpha + 2(\sqrt{q^2 + (1 - q^2)\alpha} - q)/(1 - q)\}.$$

The sets $R_{\mathcal{U}}$ and B_1 have some interesting properties. We denote the projection of the upper boundary of $R_{\mathcal{U}}$ and B_1 onto the (α, β) plane for a fixed q by $R_{\mathcal{U}}^q$ and B_1^q , respectively. It is easy to see that if $q_1 \leq q_2$, then $R_{\mathcal{U}}^{q_1}$ lies above $R_{\mathcal{U}}^{q_2}$ in the (α, β) plane. Similarly, if $q_1 \leq q_2$, then $B_1^{q_1}$ lies above $B_1^{q_2}$ in the (α, β) plane. Moreover, we have the following proposition.

Proposition 3.1.1. $B_1 \subset R_{\mathcal{U}}$.

Proof. B_1 and $R_{\mathcal{U}}$ share the same lower boundary $\beta = \alpha$. Thus, it suffices to prove that the upper boundary of B_1 is below that of $R_{\mathcal{U}}$ for a fixed q , i.e., $R_{\mathcal{U}}^q$ lies above B_1^q in the (α, β) plane.

The upper boundary of B_1 is

$$\beta = \frac{-q + \sqrt{q^2 + (1 - q)^2(2\alpha - \alpha^2) + 2q(1 - q)\alpha}}{1 - q}.$$

The upper boundary of $R_{\mathcal{U}}$ is

$$\beta = -\alpha + 2\frac{\sqrt{q^2 + (1 - q^2)\alpha} - q}{1 - q}.$$

Notice that when $q = 0$, these boundaries reduce to the boundaries in [33]. We need to prove the following:

$$\begin{aligned} & \frac{-q + \sqrt{q^2 + (1 - q)^2(2\alpha - \alpha^2) + 2q(1 - q)\alpha}}{1 - q} \\ & \leq -\alpha + 2\frac{\sqrt{q^2 + (1 - q^2)\alpha} - q}{1 - q}. \end{aligned}$$

It suffices to show that

$$\begin{aligned} & \sqrt{q^2 + (1 - q)^2(2\alpha - \alpha^2) + 2q(1 - q)\alpha} \\ & \leq -\alpha(1 - q) - q + 2\sqrt{q^2 + (1 - q^2)\alpha}. \end{aligned}$$

Squaring both sides and simplifying, we have

$$\begin{aligned} & 2\sqrt{q^2 + (1 - q^2)\alpha}(\alpha(1 - q) + q) \\ & \leq 2(q^2 + (1 - q^2)\alpha) - (1 - q)^2(\alpha - \alpha^2). \end{aligned}$$

Again squaring both sides and simplifying, we have

$$\begin{aligned}
& 4(q^2 + (1 - q^2)\alpha)(q^2 + 2q(1 - q)\alpha + (1 - q)^2\alpha^2 \\
& - q^2 - (1 - q^2)\alpha + (1 - q)^2(\alpha - \alpha^2)) \\
& \leq (1 - q)^4(\alpha - \alpha^2)^2.
\end{aligned}$$

Fortuitously, the left hand side turns out to be identically 0. Thus, the inequality holds. The reader can refer to Fig. 3.3(a) and Fig. 3.3(b) for plots of the upper boundaries of $R_{\mathcal{U}}$ and B_1 for two fixed values of q .

□

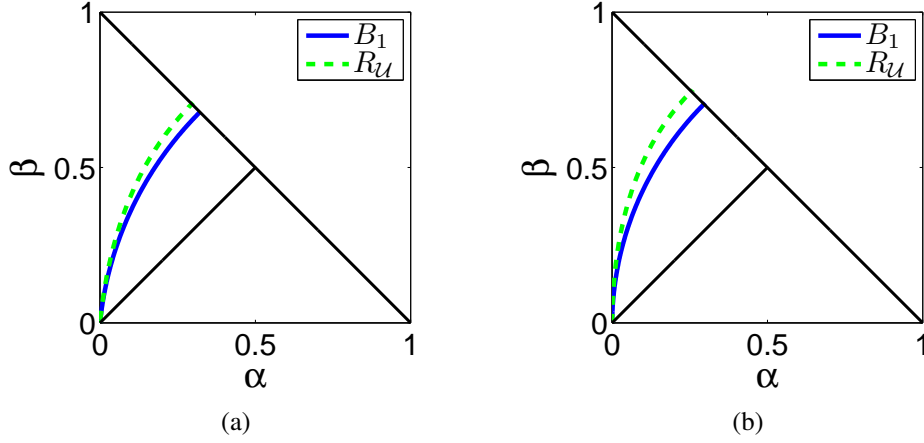


Figure 3.3: (a) Upper boundaries for $R_{\mathcal{U}}$ and B_1 for $q = 0.1$. (b) Upper boundaries for $R_{\mathcal{U}}$ and B_1 for $q = 0.01$.

We denote the region $R_{\mathcal{U}} \cup R_{\mathcal{L}}$ by R . We show below that R is an *invariant region* in the sense that once the system enters R , it stays there.

Proposition 3.1.2. *If $(\alpha_{k_0}, \beta_{k_0}, q_{k_0}) \in R$ for some k_0 , then $(\alpha_k, \beta_k, q_k) \in R$ for all $k \geq k_0$.*

Proof. Without loss of generality, we assume $(\alpha_k, \beta_k, q_k) \in R_{\mathcal{U}}$. We know that $R_{\mathcal{L}}$ is the image of \mathcal{U} in \mathcal{L} . Thus if the next state $(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) \in \mathcal{L}$, then it must be inside $R_{\mathcal{L}}$. We already have $q_{k+1} \leq q_k$, which indicates that $R_{\mathcal{U}}^{q_{k+1}}$ lies above $R_{\mathcal{U}}^{q_k}$ in the (α, β) plane. Moreover, for a fixed q , the upper boundary $R_{\mathcal{U}}^q$ is monotone increasing in the (α, β) plane. We already know that $\alpha_{k+1} > \alpha_k$ and $\beta_{k+1} < \beta_k$. As a result, if the next state $(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) \in \mathcal{U}$, then the next state is in fact inside $R_{\mathcal{U}}$.

□

We have shown that the system enters B_1 after certain levels of fusion. By the fact that $B_1 \subset R_{\mathcal{U}}$, we conclude that the system enters $R_{\mathcal{U}}$ at some level of the tree and stays inside the invariant region R at all levels above.

In the next section, we will consider the step-wise reduction of the total error probability when the system lies inside the invariant region and deduce upper and lower bounds for the total error probability.

3.1.3 Error probability bounds

Recall that the total detection error probability for a node at the k -th level is $(\alpha_k + \beta_k)/2$ because of the equal-prior assumption. Let $L_k = \alpha_k + \beta_k$, which is twice the total error probability. We will derive bounds on $\log L_k^{-1}$, whose growth rate is related to the rate of converge of L_k to 0.

Proposition 3.1.3. *Let $L_{k+1}^{(q)}$ be the total error probability at the next level from the current state (α_k, β_k, q) . Suppose that (α_k, β_k, q_1) and $(\alpha_k, \beta_k, q_2) \in \mathcal{U}$. If $q_1 < q_2$, then*

$$L_{k+1}^{(q_1)} \leq L_{k+1}^{(q_2)},$$

with equality if and only if $\alpha_k = \beta_k$.

Proof. From (3.5), we have

$$L_{k+1}^{(q)} = \frac{1-q}{1+q} L_{k+1}^{(0)} + \frac{2q}{1+q} (\alpha_k + \beta_k),$$

where $L_{k+1}^{(0)} = 2\alpha_k - \alpha_k^2 + \beta_k^2$.

It is easy to show that $2\alpha_k - \alpha_k^2 + \beta_k^2 \leq \alpha_k + \beta_k$.

$$\begin{aligned} 2\alpha_k - \alpha_k^2 + \beta_k^2 &\leq \alpha_k + \beta_k \\ \iff \alpha_k - \alpha_k^2 &\leq \beta_k - \beta_k^2. \end{aligned}$$

Since $\alpha_k + \beta_k \leq 1$ and $\beta_k \geq \alpha_k$, we have $\beta_k - 1/2 \leq 1/2 - \alpha_k$. Notice that the function $x - x^2$ peaks at $x = 1/2$. Hence, $2\alpha_k - \alpha_k^2 + \beta_k^2 \leq \alpha_k + \beta_k$ with equality if and only if $\alpha_k = \beta_k$.

Notice that

$$\frac{1-q}{1+q} + \frac{2q}{1+q} = 1.$$

Therefore we can write

$$L_{k+1}^{(q_1)} = p_1 L_{k+1}^{(0)} + (1-p_1)(\alpha_k + \beta_k),$$

where $p_1 = (1-q_1)/(1+q_1)$. Let $p_2 = (1-q_2)/(1+q_2)$, it is easy to see that $p_1 \geq p_2$. Thus we have

$$\begin{aligned} L_{k+1}^{(q_1)} &= p_1 L_{k+1}^{(0)} + (1-p_1)(\alpha_k + \beta_k) \\ &\quad + (p_2 - p_1) L_{k+1}^{(0)} - (p_2 - p_1) L_{k+1}^{(0)} \\ &\leq p_1 L_{k+1}^{(0)} + (1-p_1)(\alpha_k + \beta_k) \\ &\quad + (p_2 - p_1) L_{k+1}^{(0)} - (p_2 - p_1)(\alpha_k + \beta_k) \\ &= L_{k+1}^{(q_2)}. \end{aligned}$$

Moreover, ‘=’ holds if and only if $\alpha_k = \beta_k$. □

From Proposition 3.1.3, we immediately deduce that for any $q_1 > 0$,

$$L_{k+1}^{(0)} \leq L_{k+1}^{(q_1)}.$$

This means that the decay of the total error probability for a single step is the fastest when $q = 0$. As a result, for the case where $q \neq 0$, the step-wise shrinkage of the total error probability cannot be faster than the case where $q = 0$, where the asymptotic decay exponent is \sqrt{N} [33].

Notice that from (3.1), the decay of q_k is quadratic, which is much faster than the decay rate of L_k . Moreover, it is easy to see that the decay of q_k is faster than the decay of α_k and of β_k . Hence, it is natural to assume that $q_k \leq \alpha_k$ and $q_k \leq \beta_k$ when we consider the step-wise shrinkage of the total error probability in the invariant region. Next we give upper and lower bounds for the ratio L_{k+2}/L_k^2 .

Proposition 3.1.4. *Suppose that $(\alpha_k, \beta_k, q_k) \in R$, $\alpha_k \geq q_k$, and $\beta_k \geq q_k$. Then,*

$$\frac{1}{2} \leq \frac{L_{k+2}}{L_k^2} \leq 4.$$

Proof. First, we consider the lower bound. The evolution of the system is

$$(\alpha_k, \beta_k, q_k) \rightarrow (\alpha_{k+1}, \beta_{k+1}, q_k^2) \rightarrow (\alpha_{k+2}, \beta_{k+2}, q_k^4).$$

From Proposition 3.1.3, we have

$$L_{k+2}^{(0)} \leq L_{k+2},$$

where $L_{k+2}^{(0)} = 2\alpha_{k+1} - \alpha_{k+1}^2 + \beta_{k+1}^2$ as defined before. To prove $1/2 \leq L_{k+2}/L_k^2$, it suffices to show that $1/2 \leq L_{k+2}^{(0)}/L_k^2$.

If $(\alpha_k, \beta_k) \in R_u \setminus B_1$, then

$$\frac{L_{k+2}^{(0)}}{L_k^2} = \frac{2\alpha_{k+1} - \alpha_{k+1}^2 + \beta_{k+1}^2}{(\alpha_k + \beta_k)^2}.$$

We have

$$\alpha_{k+1} = \frac{1 - q_k}{1 + q_k}(2\alpha_k - \alpha_k^2) + \frac{2q_k}{1 + q_k}\alpha_k \geq \alpha_k$$

and

$$\beta_{k+1} = \frac{1 - q_k}{1 + q_k}\beta_k^2 + \frac{2q_k}{1 + q_k}\beta_k \geq \beta_k^2.$$

Thus, it suffices to show that

$$\frac{2\alpha_k - \alpha_k^2 + \beta_k^4}{(\alpha_k + \beta_k)^2} \geq \frac{1}{2}.$$

It is easy to see that

$$2(2\alpha_k - \alpha_k^2) \geq 1 - (1 - \alpha_k)^4.$$

Hence, it suffices to show that

$$(1 - (1 - \alpha_k)^4 + \beta_k^4) \geq (\alpha_k + \beta_k)^2,$$

which has been proved in [33].

If $(\alpha_k, \beta_k) \in B_1$, then it suffices to show that

$$\frac{\alpha_{k+1}^2 + 2\beta_{k+1} - \beta_{k+1}^2}{(\alpha_k + \beta_k)^2} \geq \frac{1}{2}.$$

We have

$$\alpha_{k+1} = \frac{1 - q_k}{1 + q_k}(2\alpha_k - \alpha_k^2) + \frac{2q_k}{1 + q_k}\alpha_k \geq \alpha_k$$

and

$$\beta_{k+1} = \frac{1 - q_k}{1 + q_k}\beta_k^2 + \frac{2q_k}{1 + q_k}\beta_k \geq \beta_k^2.$$

Thus, it suffices to proof

$$\frac{\alpha_k^2 + \beta_k^2}{(\alpha_k + \beta_k)^2} \geq \frac{1}{2},$$

which is easy to see.

Next we prove the upper bound of the ratio L_{k+2}/L_k^2 .

If $(\alpha_k, \beta_k) \in R_u \setminus B_1$, then

$$\frac{L_{k+2}}{L_k^2} \leq \frac{L_{k+1}}{L_k^2} = \frac{1 - q_k}{1 + q_k} \frac{2\alpha_k - \alpha_k^2 + \beta_k^2}{(\alpha_k + \beta_k)^2} + \frac{2q_k}{1 + q_k}(\alpha_k + \beta_k).$$

It is easy to see that

$$\frac{2q_k}{1 + q_k}(\alpha_k + \beta_k) \leq 1.$$

Next, we can prove that

$$\frac{2\alpha_k - \alpha_k^2 + \beta_k^2}{(\alpha_k + \beta_k)^2} \leq 2,$$

which is equivalent to

$$\phi(\alpha_k, \beta_k) := 2\alpha_k - 3\alpha_k^2 - \beta_k^2 - 4\alpha_k\beta_k \leq 0.$$

We have

$$\frac{\partial \phi}{\partial \beta_k} = -2\beta_k - 4\alpha_k \leq 0.$$

Thus, we can consider the lower boundary of this region which is the upper boundary of B_1 .

$$\beta = \frac{-q + \sqrt{q^2 + (1 - q)^2(2\alpha - \alpha^2) + 2q(1 - q)\alpha}}{1 - q}.$$

Denote $\varphi(\alpha, q) := \sqrt{q^2 + (1 - q)^2(2\alpha - \alpha^2) + 2q(1 - q)\alpha}$. We have

$$\begin{aligned} \phi(\alpha_k, \beta_k) &= -(q_k^2 + \beta_k^2 + (1 - q_k)^2(2\alpha_k - \alpha_k^2) \\ &\quad + 2q_k(1 - q_k)\alpha_k - 2q_k\varphi(\alpha_k, q_k))/(1 - q_k)^2 \\ &\quad - 4\alpha_k\beta_k + 2\alpha_k - 3\alpha_k^2 \\ &= \frac{2q_k\beta_k}{1 - q_k} - 4\alpha_k\beta_k - \frac{2q_k\alpha_k}{1 - q_k} - 2\alpha_k^2. \end{aligned}$$

It is easy to see that

$$\frac{2q_k\beta_k}{1-q_k} - 4\alpha_k\beta_k \leq 0.$$

Hence, we have

$$\frac{1-q_k}{1+q_k} \frac{2\alpha_k - \alpha_k^2 + \beta_k^2}{(\alpha_k + \beta_k)^2} \leq 2,$$

and

$$\frac{L_{k+2}}{L_k^2} \leq 3.$$

For the case where $(\alpha_k, \beta_k) \in B_1$, we prove that the ratio is upper bounded by 4. The evolution of the system is

$$(\alpha_k, \beta_k, q_k) \rightarrow (\alpha_{k+1}, \beta_{k+1}, q_k^2) \rightarrow (\alpha_{k+2}, \beta_{k+2}, q_k^4).$$

It is easy to see that

$$L_{k+2}^{(q_k)} \geq L_{k+2},$$

where $L_{k+2}^{(q_k)}$ denotes the total error probability if we use q_k to calculate from L_{k+1} to L_{k+2} . Therefore, it suffices to prove that

$$L_{k+2}^{(q_k)} - 4L_k^2 = \alpha_{k+2} + \beta_{k+2} - 4(\alpha_k + \beta_k)^2 \leq 0.$$

We have

$$\beta_{k+1} = \frac{1-q_k}{1+q_k} \beta_k^2 + \frac{2q_k}{1+q_k} \beta_k.$$

From the assumption that $\beta_k \geq q$, we have

$$\frac{\partial \beta_{k+1}}{\partial \beta_k} = \frac{2(1-q_k)}{1+q_k} \beta_k + \frac{2q_k}{1+q_k} \leq 4\beta_k.$$

It is easy to get that

$$\begin{aligned} \beta_{k+2} &= \frac{1-q_k}{1+q_k} (2\beta_{k+1} - \beta_{k+1}^2) + \frac{2q_k}{1+q_k} \beta_{k+1} \\ &= -\frac{1-q_k}{1+q_k} \beta_{k+1}^2 + \frac{2}{1+q_k} \beta_{k+1}. \end{aligned}$$

Therefore, we have

$$\frac{\partial \beta_{k+2}}{\partial \beta_k} = -2\frac{1-q_k}{1+q_k} \beta_{k+1} \frac{\partial \beta_{k+1}}{\partial \beta_k} + \frac{2}{1+q_k} \frac{\partial \beta_{k+1}}{\partial \beta_k} \leq 8\beta_k.$$

Thus,

$$\frac{\partial L_{k+2}^{(q_k)} - 4L_k^2}{\partial \beta_k} \leq 8\beta_k - 8\alpha_k - 8\beta_k \leq 0.$$

Therefore, we can consider the lower boundary of B_1 , $\beta_k = \alpha_k$. We have

$$\begin{aligned} L_{k+2}^{(q_k)} - 4L_k^2 &= \frac{4(1-q_k)^2(1-q_k)}{(1+q_k)^3} \alpha_k^2 - 4\frac{(1-q_k)^2}{(1+q_k)^2} \alpha_k^3 \\ &\quad + \frac{2(1-q_k)^2}{(1+q_k)^2} \alpha_k^2 + \frac{8q_k}{(1+q_k)^2} \alpha_k - 16\alpha_k^2 \leq 0, \end{aligned}$$

which holds in region B_1 . Hence, the ratio is upper bounded by 4 in this region.

□

Proposition 3.1.4 gives rise to bounds on the change in the total error probability every two steps: $L_{k+2} \leq 4L_k^2$ and $L_{k+2} \geq L_k^2/2$. From these, we can derive bounds for $\log L_k^{-1}$ for even-height trees, i.e., $k = \log N$ is even. Let $P_N = L_{\log N}$, namely, the total error probability at the fusion center. We will derive bounds for $\log P_N^{-1}$.

Theorem 3.1.1. *Suppose that $(\alpha_0, \beta_0, q_0) \in R$ and $\log N$ is even. Then,*

$$\sqrt{N} (\log L_0^{-1} - 2) \leq \log P_N^{-1} \leq \sqrt{N} (\log L_0^{-1} + 1).$$

Proof. If $(\alpha_0, \beta_0, q_0) \in R$, then we have $(\alpha_k, \beta_k, q_k) \in R$ for $k = 0, 1, \dots, \log N - 2$. From Proposition 3.1.4, we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 1, \dots, \log N - 2$ and some $a_k \in [1/2, 4]$. Therefore, for $k = 2, 4, \dots, \log N$, we have

$$L_k = a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \cdots a_0^{2^{(k-2)/2}} L_0^{2^{k/2}},$$

where $a_i \in [1/2, 4]$. Substituting $k = \log N$, we have

$$\begin{aligned} \log P_N^{-1} &= -\log a_{(k-2)/2} - 2 \log a_{(k-4)/2} - \cdots \\ &\quad - 2^{(k-2)/2} \log a_0 + \sqrt{N} \log L_0^{-1}. \end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and for each i , $-1 \leq \log a_i \leq 2$. Thus,

$$\begin{aligned} \log P_N^{-1} &\leq \sqrt{N} \log L_0^{-1} + \sqrt{N} \\ &= \sqrt{N} (\log L_0^{-1} + 1). \end{aligned}$$

Finally,

$$\begin{aligned} \log P_N^{-1} &\geq -2\sqrt{N} + \sqrt{N} \log L_0^{-1} \\ &= \sqrt{N} (\log L_0^{-1} - 2). \end{aligned}$$

□

For odd-height trees, we need to calculate the decrease in the total error probability in a single step. For this, we have the following proposition.

Proposition 3.1.5. *If $(\alpha_k, \beta_k, q_k) \in \mathcal{U}$, then we have*

$$\frac{L_{k+1}}{L_k^2} \geq 1$$

and

$$\frac{L_{k+1}}{L_k} \leq 1.$$

Proof. To prove $L_{k+1}/L_k^2 \geq 1$, it suffices to prove that

$$\begin{aligned} & \frac{1-q_k}{1+q_k} (2\alpha_k - \alpha_k^2 + \beta_k^2 - (\alpha_k + \beta_k)^2) \\ & + \frac{2q_k}{1+q_k} (\alpha_k + \beta_k - (\alpha_k + \beta_k)^2) \geq 0, \end{aligned}$$

which is easy to see.

To prove $L_{k+1}/L_k \leq 1$, it suffices to prove that

$$\begin{aligned} & \frac{1-q_k}{1+q_k} (2\alpha_k - \alpha_k^2 + \beta_k^2 - (\alpha_k + \beta_k)) \\ & + \frac{2q_k}{1+q_k} (\alpha_k + \beta_k - (\alpha_k + \beta_k)) \leq 0, \end{aligned}$$

which is easy to see. □

From Propositions 3.1.4 and 3.1.5, we give bounds for the total error probability at the fusion center for trees with odd height.

Theorem 3.1.2. *Suppose that $(\alpha_0, \beta_0, q_0) \in R$ and $\log N$ is odd. Then,*

$$\sqrt{\frac{N}{2}} (\log L_0^{-1} - 2) \leq \log P_N^{-1} \leq \sqrt{2N} (\log L_0^{-1} + 1).$$

Proof. The proof is similar to that of Theorem 3.1.1 and it is omitted. □

3.1.4 Asymptotic rates

In this section, we consider the asymptotic decay rate of the total error probability with respect to N . We compare the rate with that of balanced binary relay trees without sensor failures.

Notice that when N is very large, the sequence $\{(\alpha_k, \beta_k, q_k)\}$ enters the invariant region R at some level and stays inside afterward. Therefore the decay rate in the invariant region determines the asymptotic rate. Because our error probability bounds for odd-height trees differ from those of even-height trees by a constant term, without loss of generality, we will consider trees with even height to calculate the decay rate.

Proposition 3.1.6. *If $L_0 = \alpha_0 + \beta_0$ is fixed, then*

$$\log P_N^{-1} = \Theta(\sqrt{N}).$$

Proof. To analyze the asymptotic rate, we may assume that $L_0 < 1/2$. In this case, the bounds in Theorem 3.1.1 show that

$$\log P_N^{-1} = \Theta(\sqrt{N}).$$

□

This implies that the convergence of the total error probability is sub-exponential with decay exponent \sqrt{N} . Compared to the decay exponent for the case where $q = 0$ (no sensor failures), the asymptotic rate does not change when we have crummy sensors, even though the step-wise shrinkage for the crummy sensor case is worse.

Given $L_0 \in (0, 1)$ and $\varepsilon \in (0, 1)$, suppose that we wish to determine how many sensors we need to have so that $P_N \leq \varepsilon$. The solution is simply to find an N (e.g., the smallest) satisfying the inequality

$$\sqrt{N} (\log L_0^{-1} - 2) \geq -\log \varepsilon.$$

The smallest N grows like $\Theta((\log \varepsilon)^2)$ (cf., [33], in which the growth rate is the same, and [32], where a looser bound was derived).

3.2 Communication Link Failures

Next we consider the detection performance of balanced binary relay trees with failure-prone communication links.

3.2.1 Problem formulation

We assume that all communication links between nodes at height k and height $k + 1$ have identical failure probability ℓ_k . As a result of the communication failure, with a certain probability each node at level k in the tree does not have any data, which we denote by p_k . Assuming equal prior probabilities, we use the likelihood-ratio test [41] with unit threshold when fusing binary messages at the relay nodes and the fusion center.

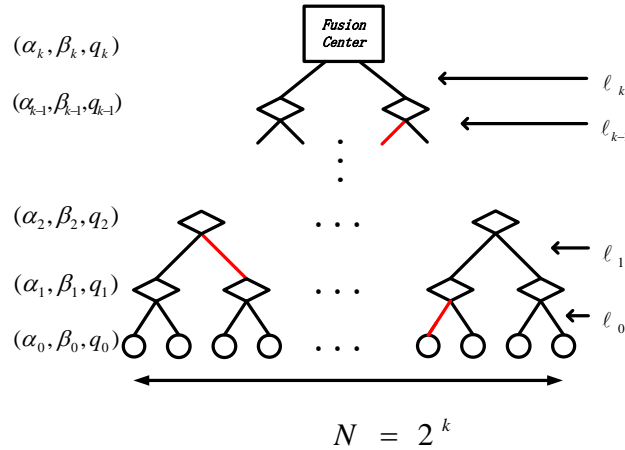


Figure 3.4: A balanced binary relay tree with height k . Circles represent sensors making measurements. Diamonds represent relay nodes which fuse binary messages. The rectangle at the root represents the fusion center making an overall decision.

Consider the simple problem of fusing binary messages passed to a node by its two immediate child nodes. Assume that the two child nodes have identical Type I error probability α , identical Type II error probability β , and identical node failure probability p . Moreover, assume that the two communication links connecting the child nodes to the parent node fail with identical probability ℓ . We can show that for each child node, the parent node will not receive any data from it with a certain probability, which we denote by q (henceforth, we call this the *node failure probability*), and given by:

$$q = p + (1 - p)\ell.$$

Denote the Type I and Type II error probability after the fusion by α' and β' . The probability that the parent node does not have any data is

$$p' = (p + (1 - p)\ell)^2 = q^2.$$

If the parent node receives data from only one of the child nodes, then the Type I and Type II error probabilities do not change since the parent node receives only one binary message. The probability of this event is $2q(1 - q)$, in which case we have

$$(\alpha', \beta') = (\alpha, \beta).$$

If the parent node receives messages from both child nodes, then the scenario is the same as that in [32] and [33]. The probability of this event is $(1 - q)^2$, in which case we have

$$(\alpha', \beta') = \begin{cases} (1 - (1 - \alpha)^2, \beta^2), & \alpha \leq \beta, \\ (\alpha^2, 1 - (1 - \beta)^2), & \alpha > \beta. \end{cases}$$

Let $\bar{\alpha}'$ and $\bar{\beta}'$ be the mean Type I and Type II error probabilities conditioned on the event that the parent node receives at least one message from its child nodes, i.e., the parent node has data. We have

$$\begin{aligned} (\bar{\alpha}', \bar{\beta}', q') &= f(\alpha, \beta, q) \\ &= \begin{cases} \left(\frac{(1-q)(2\alpha-\alpha^2)+2q\alpha}{1+q}, \frac{(1-q)\beta^2+2q\beta}{1+q}, q^2 + (1-q^2)\ell' \right), & \text{if } \alpha \leq \beta, \\ \left(\frac{(1-q)\alpha^2+2q\alpha}{1+q}, \frac{(1-q)(2\beta-\beta^2)+2q\beta}{1+q}, q^2 + (1-q^2)\ell' \right), & \text{if } \alpha > \beta. \end{cases} \end{aligned}$$

Our assumption is that all sensors have the same error probabilities (α_0, β_0, q_0) . Therefore by the above recursion, all relay nodes at level 1 will have the same *error probability triplet* $(\alpha_1, \beta_1, q_1) = f(\alpha_0, \beta_0, q_0)$ (where α_1 and β_1 are the conditional mean error probabilities). Similarly we can calculate error probability triplets for nodes at all other levels. We have

$$(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) = f(\alpha_k, \beta_k, q_k), \quad k = 1, 2, \dots, \quad (3.7)$$

where (α_k, β_k, q_k) is the error probability triplet of nodes at the k -th level of the tree.

The relation (3.7) allows us to consider (α_k, β_k, q_k) as a discrete dynamic system. For the case where $\ell_k = 0$ for all k , we have studied (see [33]) the precise evolution of the sequence $\{(\alpha_k, \beta_k)\}$, derived total error probability bounds as functions of N , and established asymptotic decay rates. In this report, we study the case where $\ell_k \neq 0$. We derive total error probability bounds and determine the decay rate of the total error probability.

We start by looking at the single trajectory shown in Fig. 3.5(a), with the communication failure probabilities given by $\ell_{k+1} = \ell_k^2$. We observe that q_k decreases very quickly to 0. In addition, as shown in Fig. 3.5(b), the trajectory approaches $\beta = \alpha$ at the beginning. After (α_k, β_k) gets too close to $\beta = \alpha$, the next pair $(\alpha_{k+1}, \beta_{k+1})$ will be repelled toward the other side of the line $\beta = \alpha$. This behavior is similar to the non-failure scenario, in which case there exists an invariant region in the sense that the system stays in the invariant region once the system enters it [33]. Is there an invariant region for the case where $q \neq 0$? We answer this question affirmatively by precisely describing this invariant region in \mathbb{R}^3 .

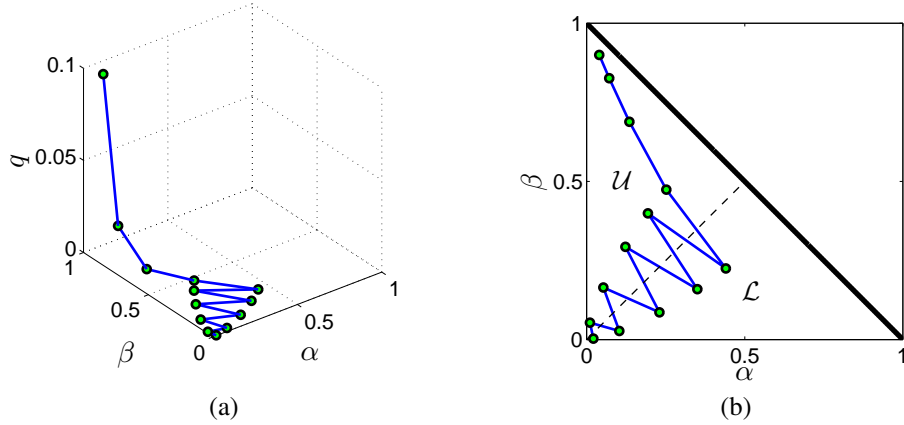


Figure 3.5: (a) A typical trajectory of (α_k, β_k, q_k) in the (α, β, q) coordinates. (b) The trajectory in (a) projected onto the (α, β) plane.

3.2.2 The evolution of Type I, Type II, and node failure probabilities

The relation (3.7) is symmetric about the hyperplanes $\alpha + \beta = 1$ and $\beta = \alpha$. Thus, it suffices to study the evolution of the dynamic system only in the region bounded by $\alpha + \beta < 1$, $\beta \geq \alpha$, and $0 \leq q \leq 1$. Let

$$\mathcal{U} := \{(\alpha, \beta) \geq 0 | \alpha + \beta < 1, \beta \geq \alpha, \text{ and } 0 \leq q \leq 1\}$$

be this triangular prism. Similarly, define the complementary triangular prism

$$\mathcal{L} := \{(\alpha, \beta) \geq 0 | \alpha + \beta < 1, \beta < \alpha, \text{ and } 0 \leq q \leq 1\}.$$

First, we introduce the following region:

$$B_1 := \{(\alpha, \beta, q) \in \mathcal{U} | \beta \leq -q/(1-q) + \sqrt{q^2 + (1-q)^2(2\alpha - \alpha^2) + 2q(1-q)\alpha}/(1-q)\}.$$

If $(\alpha_k, \beta_k, q_k) \in B_1$, then the next triplet $(\alpha_{k+1}, \beta_{k+1}, q_{k+1})$ jumps across the plane $\beta = \alpha$ away from (α_k, β_k, q_k) . More precisely, if $(\alpha_k, \beta_k, q_k) \in \mathcal{U}$, then $(\alpha_k, \beta_k, q_k) \in B_1$ if and only if $(\alpha_{k+1}, \beta_{k+1}, q_{k+1}) \in \mathcal{L}$. In other words, B_1 is the *inverse image* of \mathcal{L} in \mathcal{U} under mapping f .

It is easy to see if we start with $(\alpha_0, \beta_0, q_0) \in \mathcal{U} \setminus B_1$, then before the system enters B_1 , we have $\alpha_{k+1} > \alpha_k$ and $\beta_{k+1} < \beta_k$. Thus, the system moves towards the $\beta = \alpha$ plane. Therefore, if the number of sensors N is sufficiently large, then the system is guaranteed to enter B_1 .

Next we consider the behavior of the system after it enters B_1 . If $(\alpha_k, \beta_k, q_k) \in B_1$, we consider the position of the next pair $(\alpha_{k+1}, \beta_{k+1}, q_{k+1})$, i.e., consider the *image* of B_1 under f , denoted by $R_{\mathcal{L}}$. Similarly we denote the reflection of $R_{\mathcal{L}}$ with respect to $\beta = \alpha$ by $R_{\mathcal{U}}$. We find that

$$R_{\mathcal{U}} := \{(\alpha, \beta, q) \in \mathcal{U} | \beta \leq -\alpha + 2(\sqrt{q^2 + (1-q^2)\alpha} - q)/(1-q)\}.$$

The sets $R_{\mathcal{U}}$ and B_1 have some interesting properties. We denote the projection of the upper boundary of $R_{\mathcal{U}}$ and B_1 onto the (α, β) plane for a fixed q by $R_{\mathcal{U}}^q$ and B_1^q , respectively. It is easy to see that if $q_1 \leq q_2$, then $R_{\mathcal{U}}^{q_1}$ lies above $R_{\mathcal{U}}^{q_2}$ in the (α, β) plane. Similarly, if $q_1 \leq q_2$, then $B_1^{q_1}$ lies above $B_1^{q_2}$ in the (α, β) plane. Moreover, we have the following proposition.

Proposition 3.2.1. $B_1 \subset R_{\mathcal{U}}$.

Proof. The proof is similar to that of Proposition 3.1.1 and it is omitted. □

We denote the region $R_{\mathcal{U}} \cup R_{\mathcal{L}}$ by R . We show below that R is an *invariant region* in the sense that once the system enters R , it stays there.

Proposition 3.2.2. *If $(\alpha_{k_0}, \beta_{k_0}, q_{k_0}) \in R$ for some k_0 and $\{q_k\}$ decreases monotonically for $k \geq k_0$, then $(\alpha_k, \beta_k, q_k) \in R$ for all $k \geq k_0$.*

From the above proposition, we can study the reduction of the total error probability when the system lies in R to determine the asymptotic decay rate.

First, we compare the step-wise reduction of the total error probability between the case where communication links fail with certain probabilities (failure case) and the case where the network has no communication link failures (non-failure case). We show that if the communication links are unreliable, then the decay of the total error probability for a single step is slower than the non-failure case.

Proposition 3.2.3. *Let $L_{k+1}^{(q)}$ be (twice) the total error probability at the next level from the current state (α_k, β_k, q) . Suppose that (α_k, β_k, q_1) and $(\alpha_k, \beta_k, q_2) \in \mathcal{U}$. If $q_1 < q_2$, then*

$$L_{k+1}^{(q_1)} \leq L_{k+1}^{(q_2)}$$

with equality if and only if $\alpha_k = \beta_k$.

Proof. From the recursion, we have

$$L_{k+1}^{(q)} = \frac{1-q}{1+q} L_{k+1}^{(0)} + \frac{2q}{1+q} (\alpha_k + \beta_k),$$

where $L_{k+1}^{(0)} = 2\alpha_k - \alpha_k^2 + \beta_k^2$.

It is easy to show that $2\alpha_k - \alpha_k^2 + \beta_k^2 \leq \alpha_k + \beta_k$.

$$\begin{aligned} 2\alpha_k - \alpha_k^2 + \beta_k^2 &\leq \alpha_k + \beta_k \\ \iff \alpha_k - \alpha_k^2 &\leq \beta_k - \beta_k^2. \end{aligned}$$

Since $\alpha_k + \beta_k < 1$ and $\beta_k \geq \alpha_k$, we have $\beta_k - 1/2 < 1/2 - \alpha_k$. Notice that the function $x - x^2$ peaks at $x = 1/2$. Hence, $2\alpha_k - \alpha_k^2 + \beta_k^2 \leq \alpha_k + \beta_k$ with equality if and only if $\alpha_k = \beta_k$.

Notice that

$$\frac{1-q}{1+q} + \frac{2q}{1+q} = 1.$$

Therefore we can write

$$L_{k+1}^{(q_1)} = \pi_1 L_{k+1}^{(0)} + (1 - \pi_1)(\alpha_k + \beta_k),$$

where $\pi_1 = (1 - q_1)/(1 + q_1)$. Let $\pi_2 = (1 - q_2)/(1 + q_2)$, it is easy to see that $\pi_1 \geq \pi_2$. Thus we have

$$\begin{aligned} L_{k+1}^{(q_1)} &= \pi_1 L_{k+1}^{(0)} + (1 - \pi_1)(\alpha_k + \beta_k) \\ &\quad + (\pi_2 - \pi_1) L_{k+1}^{(0)} - (\pi_2 - \pi_1) L_{k+1}^{(0)} \\ &\leq \pi_1 L_{k+1}^{(0)} + (1 - \pi_1)(\alpha_k + \beta_k) \\ &\quad + (\pi_2 - \pi_1) L_{k+1}^{(0)} - (\pi_2 - \pi_1)(\alpha_k + \beta_k) \\ &= L_{k+1}^{(q_2)}. \end{aligned}$$

□

From Proposition 3.2.3, we immediately deduce that if $q > 0$, then

$$L_{k+1}^{(0)} \leq L_{k+1}^{(q)},$$

which means that the decay of the total error probability for a single step is the fastest if the failure probability is 0 (i.e., the non-failure case). In other words, for the failure case, the step-wise shrinkage of the total error probability cannot be faster than the non-failure case, where the total error probability decays to 0 with exponent \sqrt{N} [33].

Next we assume that the communication failure probabilities are identical at all levels, that is, $\ell_k = C$ for all k , where $C \in (0, 1)$. Denote $L_k = \alpha_k + \beta_k$ to be (twice) the total error probability for nodes at level k . Let $P_N = L_{\log N}$, which is (twice) the total error probability at the fusion center. We provide an upper bound for $\log P_N^{-1}$.

Theorem 3.2.1. *Suppose that $(\alpha_0, \beta_0, q_0) \in R$ and $\ell_k = C$ for all k , where $C \in (0, 1)$. Then,*

$$\log P_N^{-1} \leq \sqrt{N} (\log L_0^{-1} + 1).$$

Theorem 3.2.1 provides an upper bound for $\log P_N^{-1}$. Moreover, we can show that in the asymptotic regime,

$$\log P_N^{-1} = o(\sqrt{N}).$$

This implies that the convergence rate is strictly slower than \sqrt{N} (note that the convergence rate for the non-failure case is exactly \sqrt{N}).

Finally, we assume that the failure probabilities decay quadratically to 0, that is, $\ell_{k+1} = \ell_k^2$, where $k = 0, 1, \dots, \log N - 1$. In consequence, if α_0, β_0 , and q_0 are fixed, then we have $q_k \leq \alpha_k$ and $q_k \leq \beta_k$ for sufficiently large k . With these, we derive upper and lower bounds for $\log P_N^{-1}$.

Proposition 3.2.4. *Suppose that $(\alpha_k, \beta_k, q_k) \in R$, $\alpha_k \geq q_k$, and $\beta_k \geq q_k$. Then,*

$$\frac{1}{2} \leq \frac{L_{k+2}}{L_k^2} \leq 4.$$

Proof. The proof is similar to that of Proposition 3.1.4 and it is omitted. □

Proposition 3.2.4 gives rise to bounds on the change in the total error probability every two steps: $L_{k+2} \leq 4L_k^2$ and $L_{k+2} \geq L_k^2/2$. From these, we can derive bounds for $\log P_N^{-1}$ for even-height trees, i.e., $k = \log N$ is even.

Theorem 3.2.2. *Suppose that $(\alpha_0, \beta_0, q_0) \in R$ and $\ell_{k+1} = \ell_k^2$, where $k = 0, 1, \dots, \log N - 1$. If $\log N$ is even, then*

$$\sqrt{N} (\log L_0^{-1} - 2) \leq \log P_N^{-1} \leq \sqrt{N} (\log L_0^{-1} + 1).$$

Proof. If $(\alpha_0, \beta_0, q_0) \in R$, then we have $(\alpha_k, \beta_k, q_k) \in R$ for $k = 0, 1, \dots, \log N - 2$. From Proposition 3.2.4, we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 1, \dots, \log N - 2$ and some $a_k \in [1/2, 4]$. Therefore, for $k = 2, 4, \dots, \log N$, we have

$$L_k = a_{(k-2)/2} \cdot a_{(k-4)/2}^2 \dots a_0^{2^{(k-2)/2}} L_0^{2^{k/2}},$$

where $a_i \in [1/2, 4]$. Substituting $k = \log N$, we have

$$\begin{aligned} \log P_N^{-1} &= -\log a_{(k-2)/2} - 2 \log a_{(k-4)/2} - \dots \\ &\quad - 2^{(k-2)/2} \log a_0 + \sqrt{N} \log L_0^{-1}. \end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and for each i , $-1 \leq \log a_i \leq 2$. Thus,

$$\begin{aligned} \log P_N^{-1} &\leq \sqrt{N} \log L_0^{-1} + \sqrt{N} \\ &= \sqrt{N} (\log L_0^{-1} + 1). \end{aligned}$$

Finally,

$$\begin{aligned}\log P_N^{-1} &\geq -2\sqrt{N} + \sqrt{N} \log L_0^{-1} \\ &= \sqrt{N} (\log L_0^{-1} - 2) .\end{aligned}$$

□

For odd-height trees, we need to calculate the decrease in the total error probability in a single step. For this, we have the following proposition.

Proposition 3.2.5. *If $(\alpha_k, \beta_k, q_k) \in \mathcal{U}$, then we have*

$$\frac{L_{k+1}}{L_k^2} \geq 1$$

and

$$\frac{L_{k+1}}{L_k} \leq 1.$$

From Propositions 3.2.4 and 3.2.5, we give bounds for the total error probability at the fusion center for trees with odd height.

Theorem 3.2.3. *Suppose that $(\alpha_0, \beta_0, q_0) \in R$ and $\ell_{k+1} = \ell_k^2$, where $k = 0, 1, \dots, \log N - 1$. If $\log N$ is odd, then*

$$\sqrt{\frac{N}{2}} (\log L_0^{-1} - 2) \leq \log P_N \leq \sqrt{2N} (\log L_0^{-1} + 1) .$$

We have derived error probability bounds for balanced binary relay trees with unreliable communication links. In the next section, we will use these bounds to study the asymptotic rate of convergence.

3.2.3 Asymptotic rates

Notice that when N is very large and $\{q_k\}$ decreases monotonically, the sequence $\{(\alpha_k, \beta_k, q_k)\}$ enters the invariant region R at some level and stays inside afterward. In consequence, the decay rate in the invariant region determines the asymptotic rate. Since the error probability bounds for odd-height trees differ from those of even-height trees simply in a constant term, without loss of generality, we will consider trees with even height to calculate the decay rate.

Proposition 3.2.6. *Suppose that $L_0 = \alpha_0 + \beta_0$ is fixed and $\{q_k\}$ decreases monotonically. If $q_k \leq \alpha_k$ and $q_k \leq \beta_k$ for sufficiently large k , then*

$$\log P_N^{-1} = \Theta(\sqrt{N}).$$

This implies that the convergence of the total error probability is sub-exponential with exponent \sqrt{N} . Compared to the exponent for the non-failure case, the scaling law of the asymptotic rate does not change when we have unreliable communications,

provided the probabilities of communication failure probabilities decay to 0 sufficiently fast, even though the step-wise shrinkage for the failure case is worse.

Given $L_0 \in (0, 1)$ and $\varepsilon \in (0, 1)$, suppose that we wish to determine how many sensors we need to have so that $P_N \leq \varepsilon$. The solution is simply to find an N (e.g., the smallest) satisfying the inequality

$$\sqrt{N} (\log L_0^{-1} - 2) \geq -\log \varepsilon.$$

The smallest N grows like $\Theta((\log \varepsilon)^2)$ (cf., [33], in which the growth rate is the same, and [32], where a looser bound was derived).

Chapter 4

Submodularity and Optimality of Fusion Rules

4.1 Problem Formulation

Consider a binary hypothesis testing problem in a balanced binary relay tree with height h . The leaf nodes, depicted as circles in Fig. 5.1, are identical sensors making independent measurements. Then the measurements are compressed into binary messages and forwarded to their parent nodes. Consider a non-leaf node p . We denote by $C(p)$ the immediate child nodes of p , which receives two binary messages $Y_c \in \{0, 1\}$ for $c \in C(p)$. Then p summarizes the two binary messages into a new binary message $Y_p \in \{0, 1\}$ using fusion rule λ^p :

$$Y_p = \lambda^p(\{Y_c : c \in C(p)\}).$$

The new message Y_p is then communicated to the parent node of p . Ultimately, the fusion center generates an overall binary decision. In balanced binary relay tree with binary message alphabet, we already know that the majority rule with random tie-breaking does not change the Type I and II error probabilities [37]. In consequence, the only meaningful rules to aggregate two binary messages in this case are simply ‘AND’ and ‘OR’ rules defined as follows:

- AND rule (denote by \mathcal{A}): the parent node decides 1 if and only if both the child nodes send 1;
- OR rule (denote by \mathcal{O}): the parent node decides 0 if and only if both the child nodes send 0.

Notice that ULRT is either the \mathcal{A} rule or the \mathcal{O} rule, depending on the values of the Type I and Type II error probabilities at a particular level of the tree. Henceforth, we choose all fusion rules in the tree from $\mathcal{Y} = \{\mathcal{A}, \mathcal{O}\}$.

We assume that all sensors are identical and independent in this balanced configuration, and that all the nodes at level k use the same fusion rule λ_k ; i.e., $\lambda^p = \lambda_k$

for all every node p at the k -th level. In this case, we have shown in [33] that all the nodes at level k have the same Type I (false alarm) and Type II (missed detection) error probabilities, which we denote by α_k and β_k respectively.

We denote by $\pi^h = (\lambda_1, \lambda_2, \dots, \lambda_h)$ a fusion strategy, where $\lambda_j \in \mathcal{Y}$ denotes the fusion rule for nodes at each level j . Let the collection of all possible fusion strategies π^h be \mathcal{Y}^h , which can be written as

$$\mathcal{Y}^h = \underbrace{\mathcal{Y} \times \mathcal{Y} \times \dots \times \mathcal{Y}}_{h \text{ times}}.$$

Note that for given initial error probability pair (α_0, β_0) , the pair (α_h, β_h) depends on the strategy π^h .

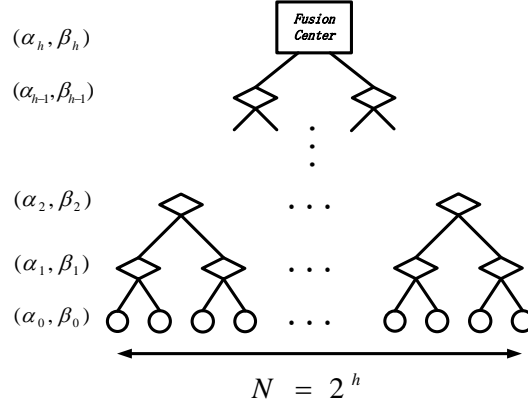


Figure 4.1: A balanced binary relay tree with height h . Circles represent sensors making measurements. Diamonds represent relay nodes which fuse binary messages. The rectangle at the root represents the fusion center making an overall decision.

For simplicity, we assume that the prior probabilities of the two hypotheses are equal. The global objective is to minimize the total error probability at the fusion center for given initial error probability pair (α_0, β_0) , namely, to maximize (twice) the reduction of the total error probability after all the fusions. We call this optimization problem a h -optimal problem, which is defined as follows:

$$\begin{aligned} v_h^*(\alpha_0, \beta_0) &= \max_{\pi^h \in \mathcal{Y}^h} (\alpha_0 + \beta_0 - (\alpha_h + \beta_h)) \\ &= \max_{\pi^h \in \mathcal{Y}^h} \sum_{j=0}^{h-1} (\alpha_j + \beta_j - (\alpha_{j+1} + \beta_{j+1})). \end{aligned}$$

The h -optimal fusion strategy is defined as

$$\begin{aligned} \pi^{h*}(\alpha_0, \beta_0) &= \arg \max_{\pi^h \in \mathcal{Y}^h} (\alpha_0 + \beta_0 - (\alpha_h + \beta_h)) \\ &= \arg \max_{\pi^h \in \mathcal{Y}^h} \sum_{j=0}^{h-1} (\alpha_j + \beta_j - (\alpha_{j+1} + \beta_{j+1})). \end{aligned}$$

On the other hand, ULRT is the 1-optimal fusion rule, which maximizes the level-wise reduction of the total error probability:

$$\text{ULRT} = \arg \max_{\lambda_i \in \mathcal{Y}} (\alpha_i + \beta_i - (\alpha_{i+1} + \beta_{i+1})) \quad \forall i.$$

Note that in this case, a maximum a posteriori fusion rule is the same to ULRT. This fusion rule is also known as the greedy rule in many literatures. In this context, we call a fusion strategy ULRT strategy if each fusion rule of the strategy is a ULRT.

In the next section, we derive the explicit fusion strategy for balanced binary relay tree with height h . We then show that the 2-optimal strategy is essentially equivalent to the ULRT strategy. Moreover, we show that the reduction of the total error probability is a submodular function, which implies that the ULRT strategy is close to optimal fusion strategy.

4.2 Main Results

4.2.1 A dynamic programming formulation

In this section, we formulate the problem using a deterministic dynamic programming model. First we define the necessary elements of this model.

- **Dynamic System:** We define the error probability pair at k -th level $s_k = (\alpha_k, \beta_k)$ to be the system state. Notice that α_k and β_k can only take values in the interval $[0, 1]$. Therefore, the set of all the states is $[0, 1] \times [0, 1]$. Moreover, given the fusion rule, the state transition function is deterministic. For example, letting $s_{k-1} = (\alpha_{k-1}, \beta_{k-1})$, if we choose $\lambda_k = \mathcal{A}$, then

$$(\alpha_k, \beta_k) := f(\alpha_{k-1}, \beta_{k-1}) = (1 - (1 - \alpha_{k-1})^2, \beta_{k-1}^2).$$

On the other hand, if we choose $\lambda_k = \mathcal{O}$, then

$$(\beta_k, \alpha_k) = f(\beta_{k-1}, \alpha_{k-1}) = (1 - (1 - \beta_{k-1})^2, \alpha_{k-1}^2).$$

- **Rewards:** At each level k , we define the instantaneous reward to be the reduction of the total error probability after fusing with λ_k :

$$r(s_{k-1}, \lambda_k) = (\alpha_{k-1} + \beta_{k-1}) - (\alpha_k + \beta_k),$$

where α_k and β_k are functions of the previous state s_{k-1} and the fusion rule λ_k .

Let $v_{h-k}(s)$ denote the cumulative reduction of the total error probability if we start the system s_k at level k and the strategy $(\lambda_{k+1}, \lambda_{k+2}, \dots, \lambda_h)$ is used. Following the above definitions, we have

$$v_{h-k}(s) = \sum_{j=k+1}^h r(s_{j-1}, \lambda_j) |_{s_k=s}.$$

If we let $k = 0$, that is, we start calculating the reduction from the sensor level, then the above cumulative reward function is the same as the global objective function defined in Section II. In consequence, for given initial state s_0 , we have to solve the following optimization problem to find the global optimal strategy over horizon h :

$$v_h^*(s_0) = \max_{\pi^h \in \mathcal{Y}^h} \sum_{j=1}^h r(s_{j-1}, \lambda_j) |_{s_0}.$$

The globally optimal strategy π^{h*} , which is a combination of fusion rules for all levels, can be written as

$$\pi^{h*}(s_0) = \arg \max_{\pi^h \in \mathcal{Y}^h} \sum_{j=1}^h r(s_{j-1}, \lambda_j) |_{s_0}.$$

Notice that the state at the k -level s_k depends on the previous state s_{k-1} and the fusion rule λ_k . Hence we write the state at level k to be $s_k |_{s_{k-1}, \lambda_k}$. The solution of the above optimization problem can be characterized using *Bellman's equation*, which states that

$$v_h^*(s_0) = \max_{\lambda_1 \in \mathcal{Y}} [r(s_0, \lambda_1) + v_{h-1}^*(s_1 |_{s_0, \lambda_1})].$$

Moreover,

$$\lambda_1^*(s_0) = \arg \max_{\lambda_1 \in \mathcal{Y}} [r(s_0, \lambda_1) + v_{h-1}^*(s_1 |_{s_0, \lambda_1})]$$

is the first element of the optimal strategy π^* . Recursively, the solution of the optimization problem is

$$v_{h-(k-1)}^*(s_{k-1}) = \max_{\lambda_k \in \mathcal{Y}} [r(s_{k-1}, \lambda_k) + v_{h-k}^*(s_k |_{s_{k-1}, \lambda_k})].$$

Moreover, the optimal fusion rule at level k is

$$\lambda_k^*(s_{k-1}) = \arg \max_{\lambda_k \in \mathcal{Y}} [r(s_{k-1}, \lambda_k) + v_{h-k}^*(s_k |_{s_{k-1}, \lambda_k})].$$

The explicit solution of the above set of equations requires dynamic programming, and the computational complexity grows exponentially with respect to the horizon.

4.2.2 2-optimal fusion strategy

In this section, we show that the 2-optimal fusion strategy is equivalent to the ULRT strategy. However, ULRT is not k -optimal for $k > 2$ in general. First consider the 2-optimal problem; i.e., $k = 2$.

$$v_2^*(s_0) = \max_{\pi^2 \in \mathcal{Y}^2} \sum_{j=1}^2 r(s_{j-1}, \lambda_j) |_{s_0},$$

where $\mathcal{Y}^2 = \{(\mathcal{A}, \mathcal{A}), (\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{O}), (\mathcal{O}, \mathcal{A})\}$. We have the following theorem.

Theorem 4.2.1. π^2 is a 2-optimal fusion strategy if and only if π^2 is a ULRT strategy.

Proof. Suppose that the initial state is (α_0, β_0) . The ULRT fusion rule reduces to \mathcal{A} and \mathcal{O} in the following way:

- ULRT= \mathcal{A} if $\beta_k \geq \alpha_k$;
- ULRT= \mathcal{O} if $\beta_k < \alpha_k$.

It is easy to show that the total error probability decreases after fusing with ULRT. However, if we apply a fusion rule other than ULRT, then we can show that the total error probability increases after fusion. For example, if $\beta_k \geq \alpha_k$ and we apply the \mathcal{O} fusion rule, then the detection error probability increases:

$$\alpha_{k+1} + \beta_{k+1} = \alpha_k^2 + 1 - (1 - \beta_k)^2 \geq \alpha_k + \beta_k,$$

In other words, the instantaneous reward is non-positive,

$$r(s_k, \mathcal{O}) \leq 0.$$

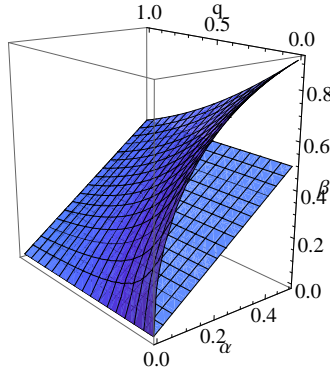


Figure 4.2: Regions \mathcal{U} , \mathcal{L} , and B_1 in the (α, β) plane.

Moreover because of symmetry, it suffices to prove this theorem in the upper triangular region $\mathcal{U} = \{(\alpha, \beta) \geq 0 | \alpha + \beta < 1 \text{ and } \beta \geq \alpha\}$ (see Fig. 4.2). Also recall that if $(\alpha_k, \beta_k) \in B_1$, where $B_1 := \{(\alpha, \beta) \in \mathcal{U} | (1 - \alpha)^2 + \beta^2 \leq 1\}$ (see Fig. 4.2), then the next state $(\alpha_{k+1}, \beta_{k+1}) \in \mathcal{L}$. Therefore, we divide our analysis into two parts:

- *Case I:* $(\alpha_0, \beta_0) \in B_1$, in which case the ULRT strategy is $(\mathcal{A}, \mathcal{O})$;
- *Case II:* $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$, in which case the ULRT fusion strategy is $(\mathcal{A}, \mathcal{A})$.

For Case I where $(\alpha_0, \beta_0) \in B_1$, it is easy to see that strategy $(\mathcal{A}, \mathcal{O})$ achieves a larger reduction than that of $(\mathcal{A}, \mathcal{A})$, because using \mathcal{A} rule for the second level increases the total error probability. Moreover, the total error probability after using $(\mathcal{O}, \mathcal{O})$ increases with respect to the initial one. Hence, this fusion rule is excluded. It suffices to show that the strategy $(\mathcal{A}, \mathcal{O})$ achieves larger reduction than that of $(\mathcal{O}, \mathcal{A})$:

$$r(s_0, \mathcal{A}) + r(s_1, \mathcal{O}) \geq r(s_0, \mathcal{O}) + r(s_1, \mathcal{A}),$$

which is equivalent with the following inequality

$$\begin{aligned} r(s_0, \mathcal{A}) + r(s_1, \mathcal{O}) - (r(s_0, \mathcal{O}) + r(s_1, \mathcal{A})) = \\ (1 - (1 - \beta_0)^2)^2 + 1 - (1 - \alpha_0^2)^2 - \\ ((1 - (1 - \alpha_0)^2)^2 + 1 - (1 - \beta_0^2)^2) \geq 0. \end{aligned}$$

The above inequality can be reduced to

$$\beta_0^2(1 - \beta_0)^2 - \alpha_0^2(1 - \alpha_0)^2 \geq 0,$$

which holds for all $(\alpha_0, \beta_0) \in B_1$.

For Case II where $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$, it is easy to see that strategy $(\mathcal{A}, \mathcal{A})$ achieves a larger reduction than that of $(\mathcal{A}, \mathcal{O})$, because using \mathcal{O} rule for the second level increases the total error probability. Moreover, the total error probability after using $(\mathcal{O}, \mathcal{O})$ increases with respect to the initial one. Hence, this fusion rule is excluded. It suffices to show that the strategy $(\mathcal{A}, \mathcal{A})$ achieves larger reduction than that of $(\mathcal{O}, \mathcal{A})$:

$$r(s_0, \mathcal{A}) + r(s_1, \mathcal{A}) \geq r(s_0, \mathcal{O}) + r(s_1, \mathcal{A}),$$

which reduces to

$$\begin{aligned} r(s_0, \mathcal{A}) + r(s_1|s_0, \mathcal{A}) - (r(s_0, \mathcal{O}) + r(s_1|s_0, \mathcal{A})) = \\ (1 - (1 - \beta_0)^2)^2 + 1 - (1 - \alpha_0^2)^2 - \\ (1 - (1 - \alpha_0)^4 + \beta_0^4) \geq 0. \end{aligned}$$

The above inequality is equivalent to

$$\beta_0(1 - \beta_0)(1 + \beta_0) - \alpha_0(1 - \alpha_0)(1 + \alpha_0) \geq 0,$$

which holds for all $(\alpha_0, \beta_0) \in \mathcal{U} \setminus B_1$.

□

We have shown that the ULRT strategy which maximizes the step-wise reduction in the total error probability is also a 2-optimal fusion strategy. However, the ULRT strategy is not in general optimal for multiple levels; i.e., $h > 2$. Next we provide a numerical example that shows that the ULRT strategy is not 3-optimal. Let the initial state be $(\alpha_0, \beta_0) = (0.2, 0.3)$. The ULRT strategy in this case is $(\mathcal{A}, \mathcal{O}, \mathcal{A})$. As shown in Fig. 4.3, the red line denotes the total error probabilities at each level up to 3. However, the 3-optimal strategy in this case is $(\mathcal{O}, \mathcal{A}, \mathcal{A})$. The total error probability curve of this strategy is shown as a green dashed-line in Fig. 4.3.

4.2.3 Submodularity

Consider a balanced binary relay tree with height $2h$. We assume that two fusion rules Λ of consecutive levels are choosing from the following set $\mathcal{Z} = \{(\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{A})\}$. Let $\Pi = (\Lambda_1, \Lambda_2, \dots, \Lambda_h)$ be the overall fusion strategy, where $\Lambda_i \in \mathcal{Z}$. In this case, the reduction of the total error probability is

$$u_h(\Lambda_1, \Lambda_2, \dots, \Lambda_h) = \alpha_0 + \beta_0 - (\alpha_{2h} + \beta_{2h}).$$

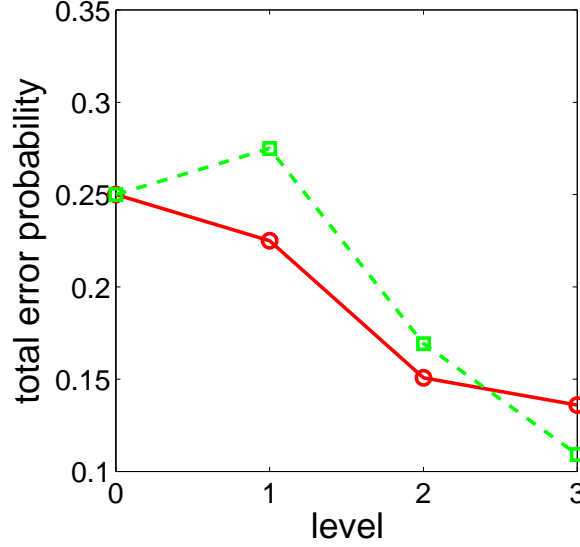


Figure 4.3: Comparison of the ULRT strategy and the 3-optimal strategy.

The global optimization problem is to select $\Lambda_i \in \mathcal{Z}$ such that the above reduction is maximized, that is

$$u_h^* = \max_{\Pi \in \mathcal{Z}^h} u_h(\Lambda_1, \Lambda_2, \dots, \Lambda_h).$$

Next we show an important property of the reduction in the total error probability.

Proposition 1: $u_h(\Lambda_1, \Lambda_2, \dots, \Lambda_h): \mathcal{Z}^h \rightarrow \mathbb{R}$ is a submodular function.

Proof. We show the ‘diminishing return’ property of the above function u_h , that is,

$$\begin{aligned} u_{m+1}(\Lambda_1, \Lambda_2, \dots, \Lambda_m, \Lambda^*) - u_m(\Lambda_1, \Lambda_2, \dots, \Lambda_m) &\geq \\ u_{n+1}(\Lambda_1, \Lambda_2, \dots, \Lambda_m, \dots, \Lambda_n, \Lambda^*) & \\ - u_n(\Lambda_1, \Lambda_2, \dots, \Lambda_m, \dots, \Lambda_n), & \end{aligned}$$

where $\Lambda_i \in \mathcal{Z}$ for all i and $\Lambda^* \in \mathcal{Z}$.

We first prove the simplest case where $m = 0$ and $n = 1$, that is,

$$u_1(\Lambda^*) - u_0(\emptyset) \geq u_2(\Lambda_1, \Lambda^*) - u_1(\Lambda_1),$$

for all $\Lambda_1, \Lambda^* \in \mathcal{Z}$. We know that $u_0(\emptyset) = 0$. Because of symmetry, it suffices to show the above inequality for the cases where $(\Lambda_1, \Lambda^*) = ((\mathcal{A}, \mathcal{O}), (\mathcal{A}, \mathcal{O}))$ and $(\Lambda_1, \Lambda^*) = ((\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{A}))$. For example, if $(\Lambda_1, \Lambda^*) = ((\mathcal{A}, \mathcal{O}), (\mathcal{A}, \mathcal{O}))$, then we can show the following

$$\alpha_{k+4} - \alpha_{k+2} - (\alpha_{k+2} - \alpha_k) = \tag{4.1}$$

$$\alpha_{k+4} + \alpha_k - 2\alpha_{k+2} = \tag{4.2}$$

$$- \alpha_k^{16} + 8\alpha_k^{14} - 24\alpha_k^{12} + 32\alpha_k^{10} \tag{4.3}$$

$$- 14\alpha_k^8 - 8\alpha_k^6 + 10\alpha_k^4 - 4\alpha_k^2 + \alpha_k \geq 0, \tag{4.4}$$

which holds for sufficiently small Type I error probability. Therefore, the inequality for the simplest case holds. The readers are referred to Fig. 4.4 for a plot of the value in Eqn. (2) with respect to α_k .

From this case, it is easy to show that

$$\begin{aligned} & u_{m+1}(\Lambda_1, \Lambda_2, \dots, \Lambda_m, \Lambda^*) - u_m(\Lambda_1, \Lambda_2, \dots, \Lambda_m) \geq \\ & u_{m+2}(\Lambda_1, \Lambda_2, \dots, \Lambda_m, \Lambda_{m+1}, \Lambda^*) \\ & - u_{m+1}(\Lambda_1, \Lambda_2, \dots, \Lambda_m, \Lambda_{m+1}), \end{aligned}$$

and

$$\begin{aligned} & u_{m+2}(\Lambda_1, \Lambda_2, \dots, \Lambda_{m+1}, \Lambda^*) - u_{m+1}(\Lambda_1, \Lambda_2, \dots, \Lambda_{m+1}) \geq \\ & u_{m+3}(\Lambda_1, \Lambda_2, \dots, \Lambda_{m+1}, \Lambda_{m+2}, \Lambda^*) \\ & - u_{m+2}(\Lambda_1, \Lambda_2, \dots, \Lambda_{m+1}, \Lambda_{m+2}), \end{aligned}$$

where $\Lambda_i \in \mathcal{Z}$ for all i and $\Lambda^* \in \mathcal{Z}$. The main result is easy to show simply by mathematical induction.

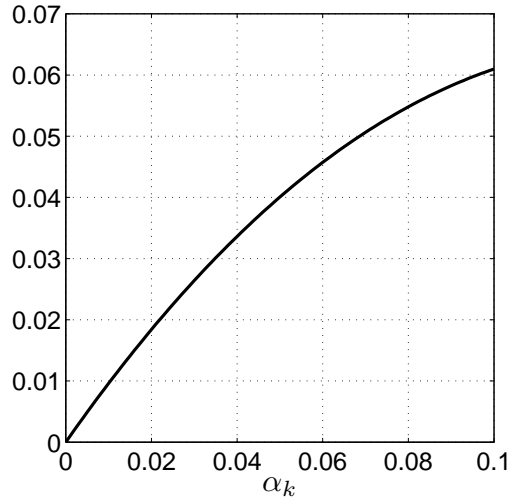


Figure 4.4: Values of $\alpha_{k+4} - \alpha_{k+2} - (\alpha_{k+2} - \alpha_k)$ versus α_k .

Next we show that the function $u_h(\Lambda_1, \Lambda_2, \dots, \Lambda_h)$ is a non-decreasing function. It suffices to show the following:

$$u_1(\Lambda^*) \geq 0,$$

for all $\Lambda^* \in \mathcal{Z}$. For example, if $\Lambda^* = (\mathcal{A}, \mathcal{O})$, then

$$u_1(\Lambda^*) = \alpha_k + \beta_k - (1 - (1 - \alpha_k)^2)^2 - (1 - (1 - \beta_k^2)^2) \geq 0,$$

which holds if and only if the Type I and II error probabilities are sufficiently small.

□

We have shown that the reduction of the total error probability is a submodular function. Moreover, we know that the total error probability does not change if there is no fusion, that is,

$$u_0(\emptyset) = 0.$$

Therefore, we can employ some well-known results about the optimality of greedy algorithms on submodular functions. First, it has been shown that the greedy algorithm provides a constant-factor approximation for the submodular problems.

Lemma 1: (Nemhauser et al., [1978]) For a monotonic submodular function F with $F(\emptyset) = 0$, the greedy strategy S^g achieves at least a constant factor of the maximum value $F^*(S)$ obtained by the globally optimal strategy; i.e.,

$$F(S^g) \geq (1 - e^{-1})F^*(S).$$

Proof. See [42] for the proof. □

We denote u^\diamond to be the reduction of the total error probability when using ULRT as the fusion rules for all levels in balanced binary relay trees. After applying Lemma 1 to our problem, we get the following theorem.

Theorem 4.2.2. *Consider a balanced binary relay tree with height $2h$. We have*

$$(1 - e^{-1})u_h^* \leq u_h^\diamond \leq u_h^*.$$

Proof. The inequality on the right hand side holds simply because u_h^* is the maximum reduction of the total error probability; i.e.,

$$u_h^\diamond \leq u_h^*.$$

We have shown that u_h is a submodular function with $u_0(\emptyset) = 0$. Therefore, we can simply apply the Lemma 1 to this problem.

$$(1 - e^{-1})u_h^* \leq u_h^\diamond.$$

□

Theorem 2 tells that the ULRT strategy is essentially close to the overall optimal strategy. However, recall that the fusion strategy is a collection of fusion rules from $\mathcal{Z} = \{(\mathcal{A}, \mathcal{O}), (\mathcal{O}, \mathcal{A})\}$. Thus, the strategies we considered in this section have at most two consecutive repeated fusion rules. For example, the strategy $(\mathcal{A}, \mathcal{A}, \mathcal{A}, \dots)$ is not considered.

Chapter 5

M -ary Relay Trees

In this chapter, we use the above approach to study the detection performance of M -ary relay trees. In contrast to the results in [36], which only address the asymptotic regime, we derive tight upper and lower bounds for the Type I and II error probabilities at the fusion center as explicit functions of N . We show that the majority dominance rule is essentially sub-optimal in the case where M is even. Specifically, our result shows that for all M ,

$$\log_2 P_N^{-1} = O(N^{\log_M \lfloor \frac{(M+1)}{2} \rfloor}),$$

a result not present in [36].

5.1 Problem Formulation

We consider the problem of binary hypothesis testing between H_0 and H_1 in an M -ary relay tree. Let \mathbb{P}_0 and \mathbb{P}_1 be the probability measures associated with the binary hypotheses. As shown in Fig. 5.1, leaf nodes are sensors undertaking initial and independent measurements of the same event. Only the leaves are sensors making measurements in this tree architecture. These measurements are compressed into binary messages and forwarded to the parent nodes at the next level. Each non-leaf node with the exception of the root, the fusion center, is a relay node, which combines M binary messages into one new binary message and forwards the new binary message to its parent node. This process takes place at each node, culminating at the fusion center at which the final decision is made based on the information received. The height of the tree is $\log_M N$, which grows as the number of sensors increases. Evidently, for $M = 2$, the structure is simply a balanced binary relay tree, which is the worst-case scenario in the sense of largest tree height among M -ary relay trees.

We assume that all sensors are independent given each hypothesis, and that all sensors have identical Type I error probability (denoted by α_0) and identical Type II error probability (denoted by β_0). We apply the majority dominance rule as the fusion rule at the relay nodes and at the fusion center. We answer the following questions about the Type I and II error probabilities:

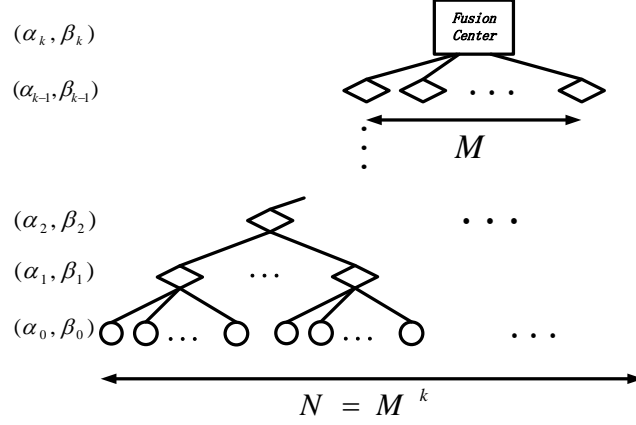


Figure 5.1: An M -ary relay tree with height k . Circles represent sensors making initial measurements. Diamonds represent relay nodes which fuse M binary messages. The rectangle at the root represents the fusion center making an overall decision.

- How do they change as we move upward in the tree?
- What are their explicit forms as explicit functions of N ?
- Do they converge to 0 at the fusion center?
- How fast will they converge with respect to N ?

5.2 Error Probability Bounds

We divide our analysis into two cases: M is an odd integer (oddary tree) and M is an even integer (evenary tree). In each case, we first derive the recursion of the Type I and II error probabilities and show that all nodes at level k have the same Type I and II error probabilities (α_k, β_k) . Then we study the step-wise reduction of each kind of error probability after fusion with majority dominance rule. From these we provide upper and lower bounds for the Type I and II error probability at the fusion center. We then derive upper and lower bounds for the total error probability at the fusion center.

5.2.1 Oddary tree

Suppose that u_o is the output binary message after fusing M input binary messages $\mathbf{u}_i = \{u_1, u_2, \dots, u_M\}$, where $u_t \in \{0, 1\}$ for all t . The majority dominance rule, when M is odd, is simply:

$$u_o := \begin{cases} 1, & \text{if } u_1 + u_2 + \dots + u_M \geq M/2, \\ 0, & \text{if } u_1 + u_2 + \dots + u_M \leq M/2. \end{cases}$$

Assume binary messages $\{u_i\}_{i=1}^M$ have identical Type I error probability α and identical Type II error probability β . Then, the Type I and II error probability pair (α', β') for the binary message u_o is given by:

$$\begin{aligned}
\alpha' &= \mathbb{P}_0(u_o = 1) = \prod_{i=1}^M \mathbb{P}_0(u_i = 1) \\
&+ \binom{M}{1} \mathbb{P}_0(u_t = 0) \prod_{i=1}^{M-1} \mathbb{P}_0(u_i = 1) + \dots \\
&+ \binom{M}{(M-1)/2} \prod_{i=1}^{(M+1)/2} \mathbb{P}_0(u_i = 1) \prod_{t=1}^{(M-1)/2} \mathbb{P}_0(u_t = 0) \\
&= f(\alpha) := \alpha^M + \binom{M}{1} \alpha^{M-1} (1 - \alpha) + \dots \\
&+ \binom{M}{(M-1)/2} \alpha^{(M+1)/2} (1 - \alpha)^{(M-1)/2}
\end{aligned}$$

and

$$\begin{aligned}
\beta' &= f(\beta) = \beta^M + \binom{M}{1} \beta^{M-1} (1 - \beta) + \dots \\
&+ \binom{M}{(M-1)/2} \beta^{(M+1)/2} (1 - \beta)^{(M-1)/2}.
\end{aligned}$$

As all sensors have the same error probability pair (α_0, β_0) , all relay nodes at level 1 will have the same error probability pair $(\alpha_1, \beta_1) = (f(\alpha_0), f(\beta_0))$. By recursion, we have

$$(\alpha_{k+1}, \beta_{k+1}) = (f(\alpha_k), f(\beta_k)), \quad k = 0, 1, \dots,$$

where (α_k, β_k) is the error probability pair of nodes at the k -th level of the tree. Since the recursions for α_k and β_k are the same, it suffices to consider only the Type I error probability α_k in studying the decay speed. Next we will analyze the step-wise shrinkage of the Type I error probability after each fusion step. This analysis will in turn provide upper and lower bounds for the Type I error probability at the fusion center.

Proposition 5.2.1. *Consider an M -ary relay tree, where M is odd. Suppose that we apply majority dominance rule as the fusion rule. Then,*

$$1 \leq \frac{\alpha_{k+1}}{\alpha_k^{(M+1)/2}} \leq 2^{M-1}.$$

Proof. Consider the ratio of α_{k+1} and $\alpha_k^{(M+1)/2}$:

$$\begin{aligned}
\frac{\alpha_{k+1}}{\alpha_k^{(M+1)/2}} &= \alpha_k^{(M-1)/2} + \binom{M}{1} \alpha_k^{(M-3)/2} (1 - \alpha_k) \\
&+ \dots + \binom{M}{(M-1)/2} (1 - \alpha_k)^{(M-1)/2}.
\end{aligned}$$

First, we derive the lower bound of the ratio. We know that

$$\begin{aligned} 1 &= (\alpha_k + 1 - \alpha_k)^{(M-1)/2} \\ &= \alpha_k^{(M-1)/2} + \binom{(M-1)/2}{1} \alpha_k^{(M-3)/2} (1 - \alpha_k) \\ &\quad + \dots + \binom{(M-1)/2}{(M-1)/2} (1 - \alpha_k)^{(M-1)/2}. \end{aligned}$$

Moreover, it is easy to see that

$$\binom{M}{k} \geq \binom{(M-1)/2}{k}$$

for all $k = 1, 2, \dots, (M-1)/2$. In consequence, we have

$$\frac{\alpha_{k+1}}{\alpha_k^{(M+1)/2}} \geq 1.$$

Next we derive the upper bound of the ratio. Since $\alpha_k < 1$, we have

$$\frac{\alpha_{k+1}}{\alpha_k^{(M+1)/2}} \leq 1 + \binom{M}{1} + \dots + \binom{M}{(M-1)/2} = 2^{M-1}.$$

□

Using the above proposition, we now derive upper and lower bounds for $\log_2 \alpha_k^{-1}$.

Theorem 5.2.1. *Consider an M -ary relay tree, where M is an odd integer. Let $\lambda_M = (M+1)/2$. We have*

$$\lambda_M^k (\log_2 \alpha_0^{-1} - (M-1)) \leq \log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}.$$

Proof. From the inequalities in Proposition 5.2.1, we have

$$\alpha_{k+1} = c_k \alpha_k^{(M+1)/2} = c_k \alpha_k^{\lambda_M},$$

where $c_k \in [1, 2^{M-1}]$. From these we obtain

$$\alpha_k = c_{k-1} c_{k-2}^{\lambda_M} \dots c_0^{\lambda_M^{k-1}} \alpha_0^{\lambda_M^k},$$

where $c_i \in [1, 2^{M-1}]$ for all i , and

$$\begin{aligned} \log_2 \alpha_k^{-1} &= -\log_2 c_{k-1} - \lambda_M \log_2 c_{k-2} - \dots \\ &\quad - \lambda_M^{k-1} \log_2 c_0 + \lambda_M^k \log_2 \alpha_0^{-1}. \end{aligned}$$

Since $\log_2 c_i \in [0, (M-1)]$, we have

$$\log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}.$$

Moreover, we obtain

$$\begin{aligned} \log_2 \alpha_k^{-1} &\geq -(M-1) - \lambda_M(M-1) - \dots \\ &\quad - \lambda_M^{k-1}(M-1) + \lambda_M^k \log_2 \alpha_0^{-1} \\ &\geq \lambda_M^k (\log_2 \alpha_0^{-1} - (M-1)). \end{aligned}$$

□

In contrast to the result in [36], which only focuses on the asymptotic regime, our result holds for all finite k . In addition, the result in [36] deals with the total error probability at the fusion center. But our approach provides bounds for both Type I and II error probabilities.

Corollary 5.2.1. *Let $P_{F,N}$ be the Type I error probability at the fusion center of an M -ary relay tree, where M is odd. We have*

$$N^{\log_M \lambda_M} (\log_2 \alpha_0^{-1} - (M-1)) \leq \log_2 P_{F,N}^{-1} \leq N^{\log_M \lambda_M} \log_2 \alpha_0^{-1}.$$

5.2.2 Evenary tree

We now study the case where M is even and derive upper and lower bounds for Type I error probability. We still use the majority dominance rule (with random tie-breaking) as the fusion rule at the relay nodes and at the fusion center. The majority dominance rule in this case is:

$$u_o := \begin{cases} 1, & \text{if } u_1 + u_2 + \dots + u_M > M/2, \\ 0 \text{ w.p. } P_b, & \text{if } u_1 + u_2 + \dots + u_M = M/2, \\ 1 \text{ w.p. } 1 - P_b, & \text{if } u_1 + u_2 + \dots + u_M = M/2, \\ 0, & \text{if } u_1 + u_2 + \dots + u_M < M/2, \end{cases}$$

where P_b denotes the Bernoulli parameter for the tie-breaking case. For simplicity, we assume that tie-breaking is fifty-fifty in this report; i.e., $P_b = 1/2$. In this case, the recursions for the Type I and II error probabilities are:

$$\begin{aligned} \alpha' &= \mathbb{P}_0(u_o = 1) = \prod_{i=1}^M \mathbb{P}_0(u_i = 1) \\ &+ \binom{M}{1} \mathbb{P}_0(u_t = 0) \prod_{i=1}^{M-1} \mathbb{P}_0(u_i = 1) + \dots \\ &+ \frac{1}{2} \binom{M}{M/2} \prod_{i=1}^{M/2} \mathbb{P}_0(u_i = 1) \prod_{t=1}^{M/2} \mathbb{P}_0(u_t = 0) \\ &= g(\alpha) := \alpha^M + \binom{M}{1} \alpha^{M-1} (1 - \alpha) + \dots \\ &+ \frac{1}{2} \binom{M}{M/2} \alpha^{M/2} (1 - \alpha)^{M/2} \end{aligned}$$

and

$$\begin{aligned} \beta' &= g(\beta) = \beta^M + \binom{M}{1} \beta^{M-1} (1 - \beta) + \dots \\ &+ \frac{1}{2} \binom{M}{M/2} \beta^{M/2} (1 - \beta)^{M/2}. \end{aligned}$$

Next we study the step-wise reduction of each type of error probability.

Proposition 5.2.2. *Consider an M -ary relay tree, where M is even. Suppose that we apply majority dominance as the fusion rule. Then,*

$$1 \leq \frac{\alpha_{k+1}}{\alpha_k^{M/2}} \leq 2^{M-1}.$$

The proof is similar to that of Proposition 5.2.1 and it is omitted. Notice that the above result is only useful when $M \geq 4$. For the case where $M = 2$ (balanced binary relay trees), we have

$$\alpha_{k+1} = \alpha_k^2 + \alpha_k(1 - \alpha_k) = \alpha_k$$

and

$$\beta_{k+1} = \beta_k^2 + \beta_k(1 - \beta_k) = \beta_k;$$

that is, the Type I and II error probabilities remain the same after fusion. However, in [33], we have shown that with the unit-threshold likelihood-ratio test as the fusion rule at the relay nodes and the fusion center, the total error probability decays to 0 sub-exponentially with exponent \sqrt{N} .

From the above proposition, we derive upper and lower bounds for the Type I error probability at each level k .

Theorem 5.2.2. *Consider an M -ary relay tree, where M is an even integer. Let $\lambda_M = M/2$. We have*

$$\lambda_M^k (\log_2 \alpha_0^{-1} - (M - 1)) \leq \log_2 \alpha_k^{-1} \leq \lambda_M^k \log_2 \alpha_0^{-1}.$$

The proof is similar to that of Theorem 5.2.1 and it is omitted. Similar with the case where M is odd, we can provide upper and lower bounds for the Type I error probability at the fusion center.

Corollary 5.2.2. *Let $P_{F,N}$ be the Type I error probability at the fusion center of an M -ary relay tree, where M is even. We have*

$$\begin{aligned} N^{\log_M \lambda_M} (\log_2 \alpha_0^{-1} - (M - 1)) &\leq \\ \log_2 P_{F,N}^{-1} &\leq N^{\log_M \lambda_M} \log_2 \alpha_0^{-1}. \end{aligned}$$

Notice that the bounds in Corollaries 5.2.1 and 5.2.2 have the same form if we simply let $\lambda_M = \lfloor (M + 1)/2 \rfloor$. In the next section, we use the bounds above to derive upper and lower bounds for the total error probability at the fusion center.

5.2.3 Bounds for the total error probability

Let π_0 and π_1 be the prior probabilities for the underlying hypotheses. In this section, we provide upper and lower bounds for the total error probability P_N at the fusion center. It is easy to see that

$$P_N = \pi_0 P_{F,N} + \pi_1 P_{M,N},$$

where $P_{F,N}$ and $P_{M,N}$ correspond to the Type I and II error probabilities at the fusion center. With the bounds for each type of error probability, we provide bounds for the total error probability as follows.

Theorem 5.2.3. Consider an M -ary relay tree, let $\lambda_m = \lfloor \frac{M+1}{2} \rfloor$. We have

$$N^{\log_M \lambda_M} (\log_2 \max\{\alpha_0, \beta_0\}^{-1} - (M-1)) \leq \log_2 P_N^{-1} \leq N^{\log_M \lambda_M} (\pi_0 \log_2 \alpha_0^{-1} + \pi_1 \log_2 \beta_0^{-1}).$$

Proof. From the definition of P_N , that is,

$$P_N = \pi_0 P_{F,N} + \pi_1 P_{M,N},$$

we have the following:

$$P_N \leq \max\{P_{F,N}, P_{M,N}\}.$$

In addition, we know that α_k and β_k have the same recursion. Therefore, the maximum between the Type I and II error probability at the fusion center corresponds to the maximum at the leaf nodes. Hence, we have

$$N^{\log_M \lambda_M} (\log_2 \max\{\alpha_0, \beta_0\}^{-1} - (M-1)) \leq \log_2 P_N^{-1}.$$

By the fact that $\log_2 x^{-1}$ is a convex function, we have

$$\log_2 P_N^{-1} \leq (\pi_0 \log_2 P_{F,N}^{-1} + \pi_1 \log_2 P_{M,N}^{-1}).$$

Therefore, we have

$$\log_2 P_N^{-1} \leq N^{\log_M \lambda_M} (\pi_0 \log_2 \alpha_0^{-1} + \pi_1 \log_2 \beta_0^{-1}).$$

□

5.2.4 Asymptotic rates

In this section, we study the decay rate of the error probabilities in the asymptotic regime. We show that in the case where M is even, the majority dominance rule is sub-optimal. We also compare our asymptotic results with those in [36].

First from Corollaries 5.2.1 and 5.2.2, we can easily derive the decay rate of the Type I and II error probabilities. For example, for the Type I error probability, we have the following.

Corollary 5.2.3. Consider an M -ary relay tree, let $\lambda_m = \lfloor \frac{M+1}{2} \rfloor$. If α_0 is fixed, then

$$\log_2 P_{F,N}^{-1} = \Theta(N^{\log_M \lambda_M}).$$

Proof. To analyze the asymptotic rate, we may assume that α_0 is sufficiently small, that is, $\alpha_0 < 2^{-(M-1)}$. In this case, the bounds in Corollaries 5.2.1 and 5.2.2 show that

$$\log_2 P_{F,N}^{-1} = \Theta(N^{\log_M \lambda_M}).$$

□

It is easy to show that $\log_M \lambda_M$ is monotone with respect to M . Moreover, as M goes to infinity, the limit of $\log_M \lambda_M$ is 1. That is to say, when M is very large, the decay is getting close to exponential. In terms of tree structures, when M is very large, the tree becomes short, and therefore achieves similar performance to that of bounded-height trees. From the fact that the Type I and II error probabilities follow the same recursion, it is easy to see that the Type II error probability at the fusion center decays to 0 with exponent $N^{\log_M \lfloor \frac{M+1}{2} \rfloor}$. Moreover, we can compute the decay rate of the total error probability.

Corollary 5.2.4. *Suppose that (α_0, β_0) is fixed. Given any prior probabilities, we have*

$$\log_2 P_N^{-1} = \Theta(N^{\log_M \lambda_M}).$$

For the total error probability at the fusion center, we have similar arguments with that for individual error probabilities. For large M , the decay of the total error probability is close to exponential.

Recall the results from [36], in which it is shown that, with any combination of fusion rules,

$$\log_2 P_N^{-1} = O(N^{\log_M \frac{(M+1)}{2}}). \quad (5.1)$$

The case where the relay nodes and the fusion center use the majority dominance rule (with random tie-breaking) to combine messages was considered in [36], in which case the decay rate of the total error probability is almost optimal. More precisely,

$$\log_2 P_N^{-1} = \Omega(N^{\log_M \lfloor \frac{(M+1)}{2} \rfloor}).$$

Our results for the odd M case is consistent with the results in [36]. The majority dominance rule in this case is essentially optimal in the sense of achieving the largest decay exponent.

$$\log_2 P_N^{-1} = \Theta(N^{\log_M \lfloor \frac{(M+1)}{2} \rfloor}) = \Theta(N^{\log_M \frac{M}{2}}). \quad (5.2)$$

However in the case where M is even, our results show that

$$\log_2 P_N^{-1} = O(N^{\log_M \lfloor \frac{(M+1)}{2} \rfloor}). \quad (5.3)$$

Compared with (5.1), which is the upper bound for $\log_2 P_N^{-1}$ in [36], our upper bound (5.3) is more tight in the case of even M and it has the exact same form with that of the lower bound; that is, we find the explicit decay rate (5.2) of the total error probability in this case. Second, the decay exponent shows that the majority dominance rule in this case is essentially sub-optimal in the sense of achieving the best decay exponent. For example, in the case of binary relay trees, the total error probability remains after fusion with the majority dominance rule. On the other hand, the likelihood-rate test with unit threshold achieves the decay exponent \sqrt{N} [33].

In [36], the case where non-binary message alphabets are allowed in M -ary relay trees is considered. Suppose that all nodes in the tree with the exception of the fusion center are allowed to transmit messages from message alphabets with size m . Then with the scheme in [36],

$$\log_2 P_N^{-1} = \Omega(N^\rho),$$

where $\rho = 1 + \ln(1 - 1/m)/\ln M$.

In this report, we introduce another message-passing scheme for M -ary relay trees with non-binary message alphabets. We show how the decay exponent increases with respect to the size of the non-binary message alphabet. Compared with the scheme in [36], we show that in order to achieve the same decay exponent, our scheme involves much lower average message sizes.

5.3 Non-binary Message Alphabets

Consider the binary hypothesis testing problem in M -ary relay trees. We characterize the detection performance by looking at the total error probability P_N at the fusion center. We have derived in [37] the decay rate of the total error probability at the fusion center in the case where relay messages are all binary, that is,

$$\log_2 P_N^{-1} = \Theta(N^{\log_M \lfloor \frac{M+1}{2} \rfloor}).$$

In this report, we allow more general message alphabet (non-binary) with size \mathcal{D} , and we denote this tree by (M, \mathcal{D}) -tree. We have studied the detection performance of $(M, 2)$ -trees in [37] by investigating how fast the total error probability decays to 0. What about the detection performance when \mathcal{D} is an arbitrary finite integer?

We denote by u_o^k the output message for each node at the k -th level after fusing M input messages $\mathbf{u}_i^{k-1} = \{u_1^{k-1}, u_2^{k-1}, \dots, u_M^{k-1}\}$ from its child nodes at the $(k-1)$ -th level, where $u_j^{k-1} \in \{0, 1, \dots, \mathcal{D}\}$ for all $j \in \{1, 2, \dots, M\}$. First, we consider an (M, \mathcal{D}) -tree with height k_0 , in which there are M^{k_0} sensors. We assume that the message alphabet size is sufficiently large; more precisely,

$$\mathcal{D} \geq M^{k_0-1} + 1. \quad (5.4)$$

Suppose that each sensor compresses its measurement into a binary message $u_o^0 \in \{0, 1\}$ and sends it upward to its parent node. Moreover, each relay node simply sums up the messages it receives from its immediate child nodes and sends the summation to its parent node; that is,

$$u_o^k = \sum_{t=1}^M u_t^{k-1}.$$

Then we can show that the output message for each node at the k -th level is an integer from $\{0, 1, \dots, M^k\}$ for all $k \in \{0, 1, \dots, k_0-1\}$. Moreover, this message essentially represents the number of its own child sensors that send ‘1’ upward. (A *child sensor* of a node in the tree is any leaf node (sensor) attached to the subtree rooted at that node.)

Because of inequality (5.4), at each level k in the tree, the message alphabet size \mathcal{D} is large enough to represent all possible values of u_o^k ($k \in \{0, \dots, k_0\}$). In particular, the fusion center (at level k_0) knows the total number of sensors that send ‘1’ upward. In this case, the detection performance is the same as a parallel configuration, where each sensor sends a binary message to the fusion center directly. Recall that in the parallel configurations, the total error probability decays exponentially fast to 0.

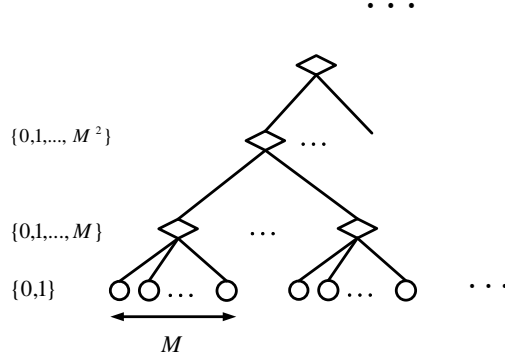


Figure 5.2: A message-passing scheme for non-binary message alphabets in M -ary relay tree.

Next we consider the case where tree height is very large. As shown in Fig. 5.2, we apply the scheme described above, that is, the sensors send binary compressions of their measurements upward to their parent nodes. Moreover, each relay node simply sends the sum of the messages received to the parent node; i.e.,

$$u_o^k = \sum_{t=1}^M u_t^{k-1}. \quad (5.5)$$

From the assumption of large tree height, it is easy to see that the message alphabet size is not large enough for all the relay nodes to use the fusion rule described in (5.5). With some abuse of notation, we let k_0 to be the integer $k_0 = \lceil \log_M(\mathcal{D} - 1) \rceil$. Note that

$$M^{k_0-1} + 1 \leq \mathcal{D} \leq M^{k_0} + 1. \quad (5.6)$$

From the previous analysis, we can show that with this scheme the nodes at the k_0 -th level knows the number of ‘1’s from its child sensors. Therefore, it is equivalent to consider the case where the nodes at level the k_0 connect to M^{k_0} sensors directly (all the intermediate relay nodes can be ignored). However, we cannot use the fusion rule described in (5.5) for the nodes at k_0 -th level to generate the output messages because the message alphabet size is not large enough. Hence, we let each node at level k_0 to aggregate the M messages from its immediate child nodes into a new binary message using the majority dominance rule (with random tie-breaking; same fusion rule as in [37]). Therefore, the output message from each node at the k_0 -th level is binary again. We can simply apply the fusion rule (5.5) and repeat this process throughout the tree, culminating at the fusion center.

Theorem 5.3.1. *The detection performance of (M, \mathcal{D}) -trees is equal to that of $(M^{k_0}, 2)$ -trees, where $k_0 = \lceil \log_M(\mathcal{D} - 1) \rceil$. In particular, if P_N be the total error probability at the fusion center for (M, \mathcal{D}) -tree, then*

$$\log_2 P_N^{-1} = \Theta(N^e),$$

where

$$\varrho := \begin{cases} \frac{\ln(M^{k_0+1})}{\ln M^{k_0}} - \frac{\log_M 2}{k_0}, & \text{if } M \text{ is odd,} \\ 1 - \frac{\log_M 2}{k_0}, & \text{if } M \text{ is even.} \end{cases}$$

Proof. Consider an (M, \mathcal{D}) -tree with the scheme described above. It is easy to see that equivalently we can consider a tree where the sensors connect to the nodes at the k_0 -th level directly. In addition, because of the recursive strategy applied throughout the tree, it suffices to consider the tree where the nodes at the ℓk_0 -th level connect to the nodes at the $(\ell+1)k_0$ -th level directly for all non-negative integers ℓ . Therefore, the detection performance of (M, \mathcal{D}) -tree is equal to that of the corresponding $(M^{k_0}, 2)$ -tree.

We have shown in [37] that the total error probability in $(M, 2)$ -trees decays to 0 with the following rate:

$$\log_2 P_N^{-1} = \Theta(N^{\log_M \lfloor \frac{M+1}{2} \rfloor}).$$

Therefore, the decay rate for $(M^{k_0}, 2)$ -trees is simply:

$$\log_2 P_N^{-1} = \Theta(N^{\log_{M^{k_0}} \lfloor \frac{M^{k_0}+1}{2} \rfloor}),$$

which can be simplified easily as follows:

$$\log_2 P_N^{-1} = \Theta(N^\varrho),$$

where

$$\varrho := \begin{cases} \frac{\ln(M^{k_0+1})}{\ln M^{k_0}} - \frac{\log_M 2}{k_0}, & \text{if } M \text{ is odd,} \\ 1 - \frac{\log_M 2}{k_0}, & \text{if } M \text{ is even.} \end{cases}$$

□

Notice that $\lim_{M \rightarrow \infty} \ln(M^{k_0} + 1) / \ln M^{k_0} = 1$, which means that the even and odd cases in the expression for ϱ are similar. Hence in the following context, we will simply analyze the case where M is even. From Theorem 5.3.1, we can see that, with larger message alphabet size, the total error probability decays more quickly. However, the change of the decay exponent is not significant because k_0 depend on \mathcal{D} logarithmically. Furthermore, if M is large, then the change of the performance becomes less sensitive to the increase in \mathcal{D} .

5.4 Scheme Comparison

In this section, we compare our scheme to that of [36]. We show that in order to achieve the same decay exponent, the average message size used in our scheme is much smaller than that used in [36].

First, notice that the result in [36] is a lower bound for the decay rate. On the other hand, our result contains the explicit decay rate of the total error probability using our scheme. The decay exponent in [36] is N^ρ , where $\rho = 1 - \log_M(m/(m-1))$. The Taylor expansion for ρ as $m \rightarrow \infty$ is

$$\rho = 1 - \frac{1}{\ln M} \left(\frac{1}{m} + \frac{1}{2m^2} + \frac{1}{3m^3} + O\left(\frac{1}{m^4}\right) \right).$$

Therefore, for large m we have

$$\rho < 1 - \frac{1}{\ln M} \left(\frac{1}{m} \right).$$

On the other hand, the decay exponent associated with our scheme is N^e , where

$$\varrho = 1 - \frac{\log_M 2}{k_0} = 1 - \frac{1}{\log_2 M} \left(\frac{1}{k_0} \right).$$

Therefore, in order to achieve the same decay exponent in the asymptotic sense, we should have

$$m = \Theta(k_0).$$

Notice that in our scheme the maximum alphabet size required is $M^{k_0-1} + 1$. Therefore, the maximum alphabet size in our scheme is much larger than that of the scheme in [36]. However, all the nodes in [36] use the same message alphabet with size m . In our scheme only very few nodes in the tree are essentially using the maximum message alphabet. For example, the sensors only send binary messages upward to their parent nodes.

It is interesting to compare the average message size used in order to achieve such detection performance. For the scheme in [36], the average message size used (in bits) is simply $b = \log_2 m = \Theta(\log_2 k_0)$. On the other hand, the average size in bits used in our scheme can be calculated as follows:

$$\bar{b}(k_0) = \frac{M^{k_0} + \dots + M \log_2(M^{k_0-1} + 1)}{M^{k_0} + M^{k_0-1} + \dots + M}$$

We have

$$\log_2(M^k + 1) > \log_2 M^k = k \log_2 M$$

and

$$\log_2(M^k + 1) < \log_2(2M^k) = 1 + k \log_2 M$$

for all k . Therefore, the average size in bits is lower bounded by the following inequality:

$$\begin{aligned} \bar{b}(k_0) &> \frac{M^{k_0} + M^{k_0-1} \log_2 M + \dots + M(k_0 - 1) \log_2 M}{M^{k_0} + M^{k_0-1} + \dots + M} \\ &= \frac{M^{k_0}}{M^{k_0} + M^{k_0-1} + \dots + M} \\ &\quad + \frac{\log_2 M (M^2(M^{k_0-1} - 1) - M(k_0 - 1)) / (M - 1)^2}{M^{k_0} + M^{k_0-1} + \dots + M} \\ &= \frac{M^{k_0} - M^{k_0-1}}{M^{k_0} - 1} + \frac{M \log_2 M}{M - 1} \frac{M^{k_0-1} - 1 - M(k_0 - 1)}{M^{k_0} - 1}. \end{aligned}$$

In addition, it is upper bounded by:

$$\bar{b}(k_0) < 1 + \frac{M \log_2 M}{M - 1} \frac{M^{k_0-1} - 1 - M(k_0 - 1)}{M^{k_0} - 1}.$$

From these inequalities, it is easy to show that as $k_0 \rightarrow \infty$,

$$\bar{b}(k_0) \nearrow 1 - \frac{1}{M} + \frac{\log_2 M}{M}.$$

Therefore, with large k_0 , the average message size in terms of bits in our scheme is much smaller than that in [36].

$$b = \log_2 m = \Theta(k_0) \gg 1 - \frac{1}{M} + \frac{\log_2 M}{M} \geq \bar{b}(k_0).$$

Chapter 6

Information Geometry of Target Tracking Sensor Networks

The work in this chapter was done by Yongqiang Cheng, Xuezhi Wang, Mark Morelande, and William Moran.

6.1 Introduction

Advanced technologies for sensing, computing and networking create enormous opportunities for handling, gathering and processing measurement information via various sensor networks. It is desirable to assess the performance of a sensor network effectively in many application fields, where the statistical properties of sensor networks are crucial. Information geometry, which is gradually gaining significance as it allows the analysis of statistical properties of sensor networks from a unified perspective, has been identified as a sophisticated and powerful tool for this purpose [43, 44].

Information geometry is the study of intrinsic properties of manifolds of probability distributions [44], where the ability of data to discriminate those distributions is translated into a Riemannian metric¹. Specifically, the Fisher information provides a local measure of discrimination of the distributions that translates immediately into a Riemannian metric on the parameter manifold of the distributions. The main tenet of information geometry is that many important notions (e.g. Fisher information, testing, estimation, estimation accuracy) in probability theory, information theory and statistics can be treated as structures (e.g. metric, divergence, projection, embedded curvatures) in differential geometry by regarding the space of probabilities as a differentiable manifold endowed with a Riemannian metric and a family of affine connections, including, but not exclusively, the canonical Levi-Civita affine connection [45]. By providing the means to analyse the Riemannian geometric properties of various families of probability density functions, information geometry offers comprehensive results about statis-

¹A Riemannian metric is an inner product defined on the tangent space of a manifold. It encodes how to measure distances, angles and area at a particular point on a manifold by specifying a scalar product between tangent vectors at that point.

tical models simply by considering them as geometrical objects.

This geometric theory of statistics was pioneered in the 1940s by Rao [46], who first interpreted the Fisher information matrix as a Riemannian metric on the space of probability distributions. Since then many scholars have contributed to the development of this theory for statistical models. In 1972, Chentsov in [47] introduced a family of affine connections and proved the uniqueness of the intrinsic metric and the one-parameter family of affine connections. Meanwhile, Efron [48] undertook pioneering work in a slightly different direction. He defined a concept of curvature called *statistical curvature* and described the basic role of curvature in the high-order asymptotic theory of statistical inference. Since then, several different groups have brought to maturity the theoretical framework of statistical geometry. Of particular note is the work of Amari and his collaborators [45, 49, 50] who have developed a duality structure theory and have unified all of these theories in a differential-geometrical framework which not only enriches the theory of information geometry but also provides opportunities for a wide range of applications. Amari’s major motivation is in machine learning. Here, we study the theory from a statistical signal processing perspective.

Information geometry has found many applications in the asymptotic theory of statistical inference [51], semiparametric statistical inference [52], the study of Boltzmann machine [53], the Expectation-Maximization (EM) algorithm [54], and learning of neural networks [55], all with certain degree of success. In the last two decades, its application has spanned several discipline areas such as information theory [56, 57], systems theory [58, 59], mathematical programming [60], and statistical physics [61, 62]. It also played a central role in the multi-terminal estimation theory [63]. In neuroscience it has been used to extract higher-order interactions among neurons [64]. Many researchers around the world are applying information geometry to new applications and formulating new interpretations. An example of the former is the derivation of the intrinsic Cramér Rao bound for the subspace tracking problem on manifolds given in [65].

Information geometry can also provide new viewpoints in the analysis of sensing systems. While important, understanding information geometry theory is nontrivial. Sensor networks for target tracking form an important class of information networks. It is well understood that the performance of target detection and tracking depends heavily on the sensing ability of the underlying sensor network, which may consist of sensors ranging from large like radars to small like motes. The advances in engineering and sensing technologies enable more complex sensor networks to be built for target detection and tracking. The evaluation of sensor network performance becomes increasingly important, in particular, for sensor network design, configuration and optimization. We believe that information geometry is able to offer advanced tools to allow us to explore and therefore understand the structures of sensor systems. This work is motivated to explore such potential in a simple and sensible way, using basic sensor problems as exemplars.

In our recent work in [66] and [67], the Integrated Fisher information distance (IFID) between two targets was approximately calculated and used to measure target resolvability in the region of interest covered by a sensor network. Nevertheless, the proposed approximation for calculating IFID is only valid for closely spaced targets and the exact IFID must be evaluated by computing the integral along the geodesic connecting the two target states, which is generally nontrivial.

In this report, the connections between information geometry and the performance of sensor networks for target tracking are explored in an attempt to gain a better understanding of sensor network measurement issues. The exact calculation of IFID and Ricci curvatures for the sensor networks with a joint likelihood are presented and analyzed. The interpretation of the geometry of statistical manifolds for sensor networks is illustrated via the affine immersion. The analysis is presented via three typical sensor network scenarios: 1) a simple range-bearing radar, 2) two bearings-only passive sonars, and 3) three ranges-only detectors, respectively. In these scenarios it is shown how information geometry can be used to address system measurement issues such as evaluating sensor capability to distinguish closely spaced targets, measuring the amount of information collected by sensors and solving the problem of optimal scheduling of network sensor and resources. Although simple synchronized sensor networks with sensors of the same type are considered in the demonstrative examples, the analysis method can be applied to a more general case where dissimilar sensors are involved as long as the likelihood and Fisher information matrix of the measurement system are available.

The major contributions of this chapter are summarised as below.

1. The IFID between the states of two targets is computed by solving the geodesic equations and is used to measure the ability of a sensor network to resolve targets. The differences between IFID and the well known Kullback-Leibler divergence are described. The relationship with the energy functional, which is the integrated differential Kullback-Leibler divergence, and the differences between it and the other two measures of divergence are described.
2. The structures of statistical manifolds are elucidated by computing the canonical Levi-Civita affine connection as well as Riemannian and scalar curvatures. The relationship between the Ricci curvature tensor field and the amount of information achievable by the network sensors is highlighted.
3. An analytical presentation of statistical manifolds as immersions in Euclidean space for the distributions of the exponential family is given.

The rest of the chapter is organized as follows. In the next section, the problem of interest and the motivations of this work are described. The principles of information geometry are then introduced in Section 6.3. In Section 6.4, sensor network information, as measured by the IFID, is analyzed for three basic types of sensor network problems; the canonical Levi-Civita affine connection as well as Riemannian and scalar curvatures are calculated to elucidate the structure of the statistical manifold; an interpretation of Ricci curvature tensor field related to information issues is discussed at the end of this section. The affine immersions of manifolds corresponding to sensor networks are presented in Section 6.5, which is followed by the conclusions in Section ??.

6.2 Target tracking in sensor networks

Let target state at time k be denoted as an n dimensional vector², i.e., $\boldsymbol{\theta}_k = [\theta_{1,k}, \dots, \theta_{n,k}]^T \in \mathbb{R}^n$, where the superscript T is the matrix transpose. Target dynamics are assumed to follow a Markov process with additive Gaussian noise.

$$\boldsymbol{\theta}_{k+1} = f(\boldsymbol{\theta}_k) + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \quad (6.1)$$

where f is the system transition (dynamical) model and \mathbf{v}_k represents process noise, which is assumed to be a zero-mean Gaussian distribution with covariance matrix \mathbf{Q}_k . The measurement of the system at time k is modeled as

$$\mathbf{x}_k = \boldsymbol{\mu}(\boldsymbol{\theta}_k) + \mathbf{w}_k, \quad \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{C}_k) \quad (6.2)$$

where $\boldsymbol{\mu}$ is the measurement-to-target state space transition function and \mathbf{w}_k is the measurement noise approximated by a zero mean Gaussian distribution with covariance matrix \mathbf{C}_k . The problem of target tracking is to find the posterior probability density of target state based on a sequence of measurements, i.e., $p(\boldsymbol{\theta}_k | \mathbf{x}_{1:k})$, where $\mathbf{x}_{1:k}$ stands for a sequence of measurements up to time k .

Standard techniques of Bayesian estimation yield a recursive solution, given by

$$p(\boldsymbol{\theta}_k | \mathbf{x}_{1:k}) = \frac{p(\mathbf{x}_k | \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k | \mathbf{x}_{1:k-1})}{\int p(\mathbf{x}_k | \boldsymbol{\theta}_k, \mathbf{x}_{1:k-1}) p(\boldsymbol{\theta}_k | \mathbf{x}_{1:k-1}) d\mathbf{x}_k} \quad (6.3)$$

where $p(\mathbf{x}_k | \boldsymbol{\theta}_k)$ is the measurement likelihood and the predicted density $p(\boldsymbol{\theta}_k | \mathbf{x}_{1:k-1})$ is determined by the posterior density $p(\boldsymbol{\theta}_{k-1} | \mathbf{x}_{1:k-1})$ at time $k-1$ and the transition density $p(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k-1})$:

$$p(\boldsymbol{\theta}_k | \mathbf{x}_{1:k-1}) = \int p(\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k-1}) p(\boldsymbol{\theta}_{k-1} | \mathbf{x}_{1:k-1}) d\boldsymbol{\theta}_{k-1}. \quad (6.4)$$

The measurement likelihood plays a central role in the ‘‘Bayesian update’’ algorithm of (6.3). In fact, this measurement density function is fully determined by the intrinsic properties of the underlying sensor network and its form has a great influence on the computational solution of the tracking problem. In practice, the likelihoods of most implementable sensor networks such as the binary sensor networks in [68] and those presented in this report belong to the exponential class of density functions, called exponential families [69]. Many popular distributions, such as Gaussian, Poisson, Gamma and Dirichet etc., are exponential families.

One of the most challenging issues in target tracking is the differentiation of a target measurement from those due to other targets and clutter, which is also known as the data association [70]. In the presence of measurement uncertainties, it is important to know how well two closely spaced targets can be differentiated using the measurements taken by a sensor network. The ability to separate two closely spaced targets using their measurements is called target resolvability, which is essentially a property of the likelihood and can be intuitively described via the platform of information geometry in terms of the IFID between two points on a statistical manifold.

²In this report, a symbol in bold face is used to denote a vector and the subscript k refers to time index. Sometimes, the time index is dropped and the subscript is subsequently used to index the location of an vector without causing confusion.

In this report, the major concern is with how much and what kind of target information can be obtained from the measurements of a sensor network at a particular time. Issues of interest include identifiability of the underlying state with respect to the sensor measurements and analysis and optimization of the information gathering capacity of the sensor network. Information geometry provides intuitive, geometrical interpretations of these problems. In the exploration of sensor network measurements in the context of statistical manifolds, the following issues are of interest:

1. Calculation of the distance i.e., the integrated Fisher information distance (IFID) between two targets on a statistical manifold and observing how the Euclidean distance differs from the information distance between two targets. The IFID is a candidate criterion to measure the resolvability of closely spaced targets. Also of interest is how the Kullback-Leibler divergence is related to the IFID. Another connected distance measure is the integrated differential Kullback-Leibler divergence, which is exactly the energy functional.
2. The amount of target information which can be acquired from the measurement of a sensor network depends on the structure of the corresponding statistical manifold, described by the curvatures of Riemannian geometry. We wish to calculate the canonical connection as well as curvatures of the statistical manifold for a given sensor network and show how the Ricci curvature tensor field is related to the performance of sensor networks in the information perspective.
3. The underlying shape of a given statistical manifold is of great interest. We explore the embedding of a parameter manifold corresponding to a given sensor network in a flat Euclidean space. Knowledge of the manifold shape is useful in the development of computational tools for measuring and controlling systems and in the optimisation of the relative error performance of using sensor network measurement.

6.3 Principles of Information Geometry

6.3.1 Definition of statistical manifold

Information geometry originates with the study of manifolds of parameters arising from parameterized families of probability distributions that are the standard basic constructs of estimation theory. Consider the parameterized family of probability distributions $S = \{p(\mathbf{x}|\boldsymbol{\theta})\}$, where \mathbf{x} is a random variable and $\boldsymbol{\theta} = [\theta_1, \dots, \theta_n]^T$ is an n dimensional parameter vector specifying the distribution. In a general context, the parameter vector resides on an abstract manifold. It is possible to think of sensing situations where the underlying manifold will have the topological structure of a more complex geometrical object such as a sphere, a torus, or the space of orthogonal matrices [65]. The family S is regarded as a statistical manifold with $\boldsymbol{\theta}$ as its (possibly local) coordinate system [64].

Figure 6.1 illustrates the definition of a statistical manifold. For a given state of interest $\boldsymbol{\theta}$ in the parameter space $\Theta \in \mathbb{R}^n$, the measurement \mathbf{x} in the sample space $\mathbf{X} \in \mathbb{R}^m$ is an instantiation of a probability distribution $p(\mathbf{x}|\boldsymbol{\theta})$. Each probability

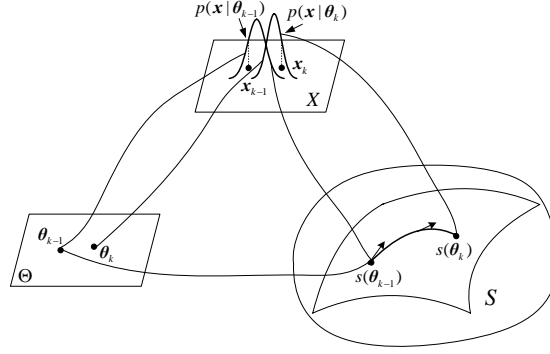


Figure 6.1: Illustration of the relation between parameter, measurement and the corresponding statistical manifold.

distribution $p(x|\theta)$ is labelled by a point $s(\theta)$ in the manifold S . The parameterized family of probability distributions $S = \{p(x|\theta)\}$ forms an n -dimensional statistical manifold where θ plays the role of a coordinate system of S . In more general situations, such as a sphere, a torus or the orthogonal group, this coordinate system will be local, and will change depending on the part of the manifold under consideration. In our examples, the coordinate system which denotes the target state of interest is global.

6.3.2 The metric and integrated fisher information distance

For a parameterized family of probability distributions on a statistical manifold, the Fisher information matrix (FIM) plays the role of a Riemannian metric tensor [46]. Denoted by $G(\theta) = [g_{ij}(\theta)]$, the FIM is defined as

$$g_{ij}(\theta) = E \left\{ \frac{\partial \log p(x|\theta)}{\partial \theta_i} \cdot \frac{\partial \log p(x|\theta)}{\partial \theta_j} \right\}, \quad (6.5)$$

where E signifies expectation. The FIM measures the ability of the random variable x to discriminate the values of the parameter θ' from θ for θ' close to θ .

The statistical manifold³ S carries the structure of a smooth Riemannian manifold whose metric is defined by the FIM $G(\theta)$ [46]. Here $\{\partial \log(\cdot)/\partial \theta_i\}$ is a basis for a vector space of random variables. The vector space is identified as the tangent space of S at θ , denoted as $\mathcal{T}_\theta S$. With this structure in place, we can bring the machinery of Riemannian geometry to bear on statistical problems. In particular, the operations of covariant differentiation can be defined to describe the various connections of interest using the one-one correspondence between the statistical parameter model and Riemannian manifold [71].

In a Riemannian manifold, the important concepts such as distance, angle and tangent are defined analogously to the case of the Euclidean space, but they only make sense locally. It is possible to integrate “distance” along curves between two points and

³In this report, the notation (S, g) is sometimes used to signify a Riemannian manifold equipped with a Fisher information metric of element g .

then select a curve of the shortest length as a distance measure between the two points on the manifold. Information geometry allows us then to establish a genuine metric between statistical distributions that is invariant to transformations by non-singular parameter transformations in [72]. To describe these ideas, we note that the infinitesimal squared distance between two closely spaced distributions $p(\mathbf{x}|\boldsymbol{\theta})$ and $p(\mathbf{x}|\boldsymbol{\theta} + d\boldsymbol{\theta})$ is given by the quadratic form of $d\boldsymbol{\theta}$ as in [64],

$$ds^2 = \sum_{ij} g_{ij} d\theta_i d\theta_j = d\boldsymbol{\theta}^T \mathbf{G}(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (6.6)$$

Consider a curve $\boldsymbol{\theta}(t) \in \boldsymbol{\Theta}$ joining $\boldsymbol{\theta}_1 = \boldsymbol{\theta}(t_1)$ and $\boldsymbol{\theta}_2 = \boldsymbol{\theta}(t_2)$, $t_1 \leq t \leq t_2$, which can be represented by a parametric equation in the parameter space with a single free parameter t , and let the distance along the curve between its endpoints, namely the two distributions $p(\mathbf{x}|\boldsymbol{\theta}_1)$ and $p(\mathbf{x}|\boldsymbol{\theta}_2)$ along $\boldsymbol{\theta}(t)$ [73], be

$$\mathcal{D}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \triangleq \int_{t_1}^{t_2} \left(\sqrt{\left(\frac{d\boldsymbol{\theta}}{dt} \right)^T \mathbf{G}(\boldsymbol{\theta}) \left(\frac{d\boldsymbol{\theta}}{dt} \right)} \right) dt \quad (6.7)$$

where \triangleq stands for “defined as”. This distance is dependent on the choice of the curve. The distance between $p(\mathbf{x}|\boldsymbol{\theta}_1)$ and $p(\mathbf{x}|\boldsymbol{\theta}_2)$ is defined as the minimum of such distances over all possible curves. The *integrated Fisher information distance* (IFID) between two distributions $p(\mathbf{x}|\boldsymbol{\theta}_1)$ and $p(\mathbf{x}|\boldsymbol{\theta}_2)$ is defined as the integral along the curve $\boldsymbol{\theta}(t)$ that minimises (6.7)[74], i.e.,

$$\mathcal{D}_F(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \triangleq \min_{\{\boldsymbol{\theta}(t): \boldsymbol{\theta}(t_1)=\boldsymbol{\theta}_1, \boldsymbol{\theta}(t_2)=\boldsymbol{\theta}_2\}} \int_{t_1}^{t_2} \left(\sqrt{\left(\frac{d\boldsymbol{\theta}}{dt} \right)^T \mathbf{G}(\boldsymbol{\theta}) \left(\frac{d\boldsymbol{\theta}}{dt} \right)} \right) dt \quad (6.8)$$

A curve on the statistical manifold which is a stationary point for $\mathcal{D}_F(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is denoted by $\gamma(t) \in S$. It is locally the shortest path joining the two points $p(\mathbf{x}|\boldsymbol{\theta}_1)$ and $p(\mathbf{x}|\boldsymbol{\theta}_2)$ on the statistical manifold and is called a *geodesic*. The existence of curves that are minimal in this sense is well documented in [44]. The IFID is a genuine metric, in the sense that it satisfies the symmetry property

$$\mathcal{D}_F(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathcal{D}_F(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1) \quad (6.9)$$

for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, and the triangle inequality

$$\mathcal{D}_F(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + \mathcal{D}_F(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \geq \mathcal{D}_F(\boldsymbol{\theta}_1, \boldsymbol{\theta}_3), \quad (6.10)$$

for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3 \in \boldsymbol{\Theta}$.

While the evaluation of the IFID is generally difficult, the distance between two distributions $p(\mathbf{x}|\boldsymbol{\theta}_1)$ and $p(\mathbf{x}|\boldsymbol{\theta}_2)$ may be approximated with a variety of alternatives. A popular alternative to the IFID is the Kullback-Leibler divergence (KLD) [74],

$$\begin{aligned} \text{KLD}[p(\mathbf{x}|\boldsymbol{\theta}_1)||p(\mathbf{x}|\boldsymbol{\theta}_2)] &= \int p(\mathbf{x}|\boldsymbol{\theta}_1) \log \frac{p(\mathbf{x}|\boldsymbol{\theta}_1)}{p(\mathbf{x}|\boldsymbol{\theta}_2)} d\mathbf{x} \\ &= E\left\{ \log p(\mathbf{x}|\boldsymbol{\theta}_1) - \log p(\mathbf{x}|\boldsymbol{\theta}_2) \right\}. \end{aligned} \quad (6.11)$$

It is also well known that the following relationship between the KLD and differential Fisher information distance holds [64, 66],

$$ds^2 = 2 \text{KLD} [p(\mathbf{x}|\boldsymbol{\theta})||p(\mathbf{x}|\boldsymbol{\theta} + d\boldsymbol{\theta})] \quad (6.12)$$

The KLD allows for the approximation of information distance in the absence of statistical manifold geometry, though it is not a genuine metric [74]; it fails to be symmetric or satisfy the triangle inequality. The failure of KLD to be a metric is a serious issue in terms of its use to measure difference between distributions. The axioms of symmetry and the triangle inequality are natural and accord with our intuitive notion of distance. A phrase like “the KLD between p_1 and p_2 ” is meaningless, for example, without explaining that it is the KLD *from* p_1 *to* p_2 . While the IFID is more difficult to calculate, it carries the same kind of information theoretic significance as the KLD while being a genuine metric.

An intermediate between the KLD and the IFID is the energy functional. It is the integral of the differential KLD. As aforementioned the KLD satisfies (6.12) for small perturbations of $\boldsymbol{\theta}$. However, the failure of KLD to be a metric means that integration of this infinitesimal quantity along a geodesic does not return the KLD between the endpoints. Thus, a different quantity, which is the integral of this differential, or up to a factor of two, can be calculated. This object, well known in differential geometry as the *energy functional* $\mathcal{E}(\boldsymbol{\theta})$, describes the total kinematic energy increment of a free particle (of unit mass) moving along a curve $\boldsymbol{\theta}$ from $\boldsymbol{\theta}(t_1)$ to $\boldsymbol{\theta}(t_2)$ in the manifold equipped with metric $\mathbf{G}(\boldsymbol{\theta})$, i.e.,

$$\mathcal{E}(\boldsymbol{\theta}) \triangleq \frac{1}{2} \int_{t_1}^{t_2} \left(\frac{d\boldsymbol{\theta}}{dt} \right)^T \mathbf{G}(\boldsymbol{\theta}) \left(\frac{d\boldsymbol{\theta}}{dt} \right) dt. \quad (6.13)$$

Surprisingly, minimizing this energy $\mathcal{E}(\boldsymbol{\theta})$ with respect to the curve $\gamma(t)$ leads to the same equations, i.e., the Euler-Lagrange equations in the local parameter $\boldsymbol{\theta}$, as the solution of a geodesic parameterized by arc length t [75]. In this report, we refer to the integral (6.13) along a geodesic path $\gamma(t)$ between the two distributions as the Energy Difference (ED) $\mathcal{E}_g(\gamma)$.

6.3.3 Geodesics and exponential map

The IFID and the ED between two distributions are defined in terms of the shortest geodesic in the Riemannian (statistical) manifold. There may be several different geodesics connecting two points such as two points on a torus. In the study of sensor networks, the trajectories of geodesics in the Euclidean space are of interest. Rigorously speaking, the definition of a geodesic as a stationary point of the distance integral on a smooth manifold S with affine connection ∇ means that the curve $\gamma(t)$ is such that parallel transport along the curve preserves the tangent vector to the curve [76]. Using local coordinates on S , the geodesic equations are given by the Euler-Lagrange equations as [71]

$$\frac{d^2\theta_k}{dt^2} + \sum_{i=1}^n \sum_{j=1}^n \Gamma_{ij}^k \frac{d\theta_i}{dt} \frac{d\theta_j}{dt} = 0, \quad \forall k \in \{1, \dots, n\} \quad (6.14)$$

where $\boldsymbol{\theta}(t) = [\theta_1(t), \dots, \theta_n(t)]^T$ are the coordinates of the curve $\gamma(t)$, Γ_{ij}^k , $i, j = 1, \dots, n$ are the *Christoffel symbols of the second kind* and are defined as Riemannian connection coefficients,

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{l=1}^n g^{kl} \left(\frac{\partial g_{il}}{\partial \theta_j} + \frac{\partial g_{jl}}{\partial \theta_i} - \frac{\partial g_{ij}}{\partial \theta_l} \right) \quad (6.15)$$

where $[g^{kl}]$ is the inverse of the FIM $\mathbf{G} = [g_{kl}]$.

The geodesic equations in (6.14) are ordinary differential equations for the coordinates $\boldsymbol{\theta}(t)$. A unique solution $\boldsymbol{\theta}(t)$ can be found for given initial conditions $\boldsymbol{\theta}(0)$ and $\dot{\boldsymbol{\theta}}$, which is analogous to an initial position $\boldsymbol{\theta}(0)$ and the “speed” $\boldsymbol{\nu} \in \mathcal{T}_{\boldsymbol{\theta}} S$ in the sense of the classical mechanics.

Assume that a geodesic is projected onto the parameter space Θ with a starting point $\boldsymbol{\theta}(0)$ and a tangent vector $\boldsymbol{\nu}$. The exponential map of the starting point is then defined as [76]

$$\exp_{\boldsymbol{\nu}} [\boldsymbol{\theta}(0)] \triangleq \boldsymbol{\Psi}(1; \boldsymbol{\theta}(0), \boldsymbol{\nu}). \quad (6.16)$$

where the notation $\boldsymbol{\Psi}(t; \boldsymbol{\theta}(0), \boldsymbol{\nu})$ is used to signify a geodesic with a starting point $\boldsymbol{\theta}(0)$, a tangent vector $\boldsymbol{\nu}$ and end point $\boldsymbol{\theta}(t)$.

It can be shown that the length along the geodesic between $\boldsymbol{\theta}(0)$ and $\boldsymbol{\Psi}(1; \boldsymbol{\theta}(0), \boldsymbol{\nu})$ is $|\boldsymbol{\nu}|$ [76, 77]. The above concepts are appealing because a geodesic connects two points in the Riemannian manifold with the minimum length. In classical mechanics, the geodesics can be thought of as trajectories of free particles in a manifold. Newton’s Laws allow one to relate the position, velocity, acceleration and various forces acting on a body and state this relation as a differential equation for the unknown position of the body as a function of time. When the motion of the body is at “constant speed”, no additional force is acting on the body and the trajectory of the body is a geodesic. Then, the distance from $\boldsymbol{\theta}(0)$ to $\boldsymbol{\theta}(t)$ along the geodesic is proportional to t , or more precisely, is equal to $|\boldsymbol{\nu}|t$. Iterative application of exponential maps therefore forms an approximation of flows along the geodesic and the optimization can converge quickly [77]. In general, obtaining the exponential map is a non-trivial task. In most cases, the partial derivatives of the Riemannian tensor in (6.15) lead to a rather complicated expression of Γ_{ij}^k which prevents solution of the differential equations (6.14).

6.3.4 Curvatures and information

In the mathematical field of differential geometry, the Riemann curvature tensor is the standard way to express curvature of Riemannian manifolds. To each point it associates a tensor that measures the extent to which the metric tensor is not locally isometric to a Euclidean space. In local coordinates, the Riemann curvature tensor components R_{ijk}^l are given by [78]:

$$R_{ijk}^l = \frac{\partial}{\partial \theta^j} \Gamma_{ik}^l - \frac{\partial}{\partial \theta^k} \Gamma_{ij}^l + \sum_s (\Gamma_{js}^l \Gamma_{ik}^s - \Gamma_{ks}^l \Gamma_{ij}^s) \quad (6.17)$$

where Γ_{ij}^k are the Christoffel symbols of the second kind and are given in (6.15), the integers $i, j, k, l \in [1, n]$ are the indices of coordinate components.

Ricci curvature tensor represents the amount by which the shape of the volume element of a geodesic ball in a curved Riemannian manifold deviates from that of the standard ball in Euclidean space. As such, it provides one way of measuring the degree to which the geometry determined by a given Riemannian metric might differ from that of ordinary Euclidean n -space. The Ricci curvature tensor is essentially the unique way of contracting the Riemann tensor [78],

$$R_{ij} = \sum_l R_{ilj} = \sum_l \left(\frac{\partial}{\partial \theta^l} \Gamma_{ij}^l - \frac{\partial}{\partial \theta^j} \Gamma_{il}^l \right) + \sum_{l,m} (\Gamma_{ij}^l \Gamma_{lm}^m - \Gamma_{il}^m \Gamma_{jm}^l) \quad (6.18)$$

In this report, the matrix form of Ricci curvature tensor is denoted by $\mathbf{R} = [R_{ij}]$.

The scalar curvature (or Ricci scalar) R is the simplest curvature invariant of a Riemannian manifold and is defined as the trace of the Ricci curvature,

$$R = \sum_{i,j} g^{ij} R_{ij} \quad (6.19)$$

where g^{ij} is the (i, j) th element of the inverse matrix of the FIM $\mathbf{G}(\boldsymbol{\theta})$. To each point on a Riemannian manifold, the scalar curvature assigns a single real number determined by the intrinsic geometry of the manifold near that point. Specifically, the scalar curvature represents the amount by which the volume of a geodesic ball in a curved Riemannian manifold deviates from that of the standard ball in Euclidean space. In two dimensions, the scalar curvature completely characterizes the curvature of a surface [78], but this fails in higher dimensions.

In the vicinity of an initial point at which an arbitrarily selected geodesic γ starts, the Ricci curvature describes the second order rate of change of the flux of geodesics initially parallel with γ [79]. This means that the Ricci curvature measures how the fluxes of initially parallel geodesics change in a given direction of interest, and therefore, it provides a measure of how well the neighboring fibers of geodesics stick together along their direction of elongation. As a special case, the fluxes of a bundle of parallel geodesics (straight lines) in Euclidean space have no change along their elongation. The behavior of a collection of geodesics reflects the value of the curvature as well as the structure of the manifold. Non-negative Ricci curvature implies stability and a relatively stable bundle of geodesics, while on the other hand negative Ricci curvature implies a less coherent bundle of geodesics.

Geometrically, the Ricci curvature tensor is the mathematical object that controls the growth rate of the volume of metric balls in a Riemannian manifold. The evolution of volumes under the geodesics with parallel initial tangent vectors near a point in the manifold is controlled by the Ricci curvature [80]. On the other hand, the Ricci curvature represents the amount by which the volume element of a geodesic ball in a manifold deviates in shape from that of the standard ball in Euclidean space. As such, it provides one way of measuring the degree to which the geometry determined by a given Riemannian metric might differ from that of ordinary Euclidean n -space. Near any point $\boldsymbol{\theta}$ in a Riemannian manifold (S, g) , the infinitesimal volume element $d\mu_g$ in local normal coordinates has the following expansion at $\boldsymbol{\theta}$ [81, 82]:

$$d\mu_g = \sqrt{|\mathbf{G}(\boldsymbol{\theta})|} d\mu_{Euclidean} \quad (6.20)$$

$$= \left[1 - \frac{1}{6} \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta} + O(|\boldsymbol{\theta}|^3) \right] d\mu_{Euclidean} \quad (6.21)$$

where $d\mu_{Euclidean}$ denotes the infinitesimal volume element in the parameter space (a Euclidean space here), \mathbf{R} is the matrix of Ricci curvature tensor given by Eq. (6.18) and $|\mathbf{A}|$ denotes the determinant of matrix \mathbf{A} .

Eq. (6.21) signifies that if the Ricci curvature is negative at θ in the manifold, the unit geodesic ball will have larger volume than it would in Euclidean space. In the statistical manifolds of sensor networks, the value of the determinant of the Fisher information matrix $|\mathbf{G}(\theta)|$ represents the amount of information that can be acquired by the underlying sensor system [83]. Therefore in this report, the Ricci scalar curvature specifically indicates the amount of information which can be collected by the sensor networks.

The above curvatures are defined using the canonical Levi-Civita connection (6.15) which is a natural connection compatible with the Fisher information metric. In fact, there are a variety of kinds of connections. In 1972, Chentsov [47] introduced a one-parameter family of affine connections called α -connections which were later popularized by Amari [84]. In particular, $\alpha = 0$ corresponds to the Levi-Civita connection (information connection), $\alpha = 1$ defines the e -connection (exponential connection) while $\alpha = -1$ defines the m -connection (mixture connection). The curvatures corresponding to these three connections are the Riemann curvature, e -curvature and m -curvature, respectively. All of these curvatures are intrinsic.

Many statisticians have attempted to show the relationship between curvature and statistics in the literature. The work pioneered by Efron [48] in 1975 introduced the concept of statistical curvature and described the basic role of the curvature in the high-order asymptotic theory of statistical inference. He proved that the second-order information loss of a first-order efficient estimator is related to the statistical curvature of a curve representing a one-parameter family of distributions. Dawid [85] and Madsen [86] succeeded in extending the result of Efron to the multi-parameter case while Amari unified all of these theories in a differential-geometrical framework in [84]. In this work, it was shown that the second-order information loss of a general Fisher efficient estimator can be decomposed into the sum of two non-negative terms. One is related to the e -curvature of the statistical model and the other is related to the m -curvature of the ancillary subspace associated with the estimator.

Nevertheless, in this report, we mainly indicate the relationship between the Riemann/Ricci curvature ($\alpha = 0$) and the amount of information which can be collected by the sensor networks. This is demonstrated in three scenarios in the following section.

6.4 Application Examples in sensor networks

In this Section, the analysis of sensor networks for target tracking via statistical manifold techniques and information geometry is demonstrated using the following three 2D sensor network examples:

- Example 1:** Sensor network contains a single range-bearing sensor;
- Example 2:** Sensor network contains two bearings-only sensors;
- Example 3:** Sensor network involves three range-only sensors.

In these examples, the state of a target is represented by the 2D target location, i.e., $\boldsymbol{\theta} = [\theta_1, \theta_2]^T = [x, y]^T$. The IFID between two targets (i.e., the geodesic length between probability distributions in the statistical manifolds) is calculated and is compared with the corresponding KLD. Using IFID as a measure of the resolvability of closely spaced targets is illustrated. In addition, the canonical Levi-Civita affine connection as well as Riemannian and scalar curvatures are computed to elucidate the structure of the statistical manifold. The relationship between the Ricci curvature tensor field and the maximum amount of information that can be obtained by the network sensors is discussed.

6.4.1 Geodesics and fisher information distance

In (6.2), it is assumed that the measurement vector \boldsymbol{x} of the underlying sensor networks obeys a multivariate normal distribution which belongs to a class of exponential family distributions⁴,

$$\boldsymbol{x}|\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{C}(\boldsymbol{\theta})). \quad (6.22)$$

Therefore, the likelihood of the measurement \boldsymbol{x} is given by

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = |2\pi\boldsymbol{C}(\boldsymbol{\theta})|^{-1/2} \exp \left\{ -[\boldsymbol{x} - \boldsymbol{\mu}(\boldsymbol{\theta})]^T \boldsymbol{C}^{-1}(\boldsymbol{\theta}) [\boldsymbol{x} - \boldsymbol{\mu}(\boldsymbol{\theta})] / 2 \right\}. \quad (6.23)$$

As shown in [67], the Fisher information matrix of this type of density with respect to $\boldsymbol{\theta}$ is of the form

$$g_{ij}(\boldsymbol{\theta}) = \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} \right]^T \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j} \right] + \frac{1}{2} \text{tr} \left[\boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{C}(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{C}(\boldsymbol{\theta})}{\partial \theta_j} \right] \quad (6.24)$$

A single range-bearing sensor network – Example 1

In this example, the sensor observes both range and bearing of a target, hence the measurement model (6.2) is written as

$$\boldsymbol{x} = \begin{bmatrix} r \\ \phi \end{bmatrix} = \begin{bmatrix} \sqrt{x^2 + y^2} \\ \arctan(y/x) \end{bmatrix} + \begin{bmatrix} w_r \\ w_\phi \end{bmatrix} \quad (6.25)$$

where r and ϕ signify the target range and bearing, respectively, relative to a sensor placed at the origin of the coordinate frame and the zero-mean additive noise $\boldsymbol{w} = [w_r, w_\phi]^T$ has covariance $\boldsymbol{C}(\boldsymbol{\theta})$. In view of (6.22) and (6.25), we have

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{bmatrix} \sqrt{x^2 + y^2} \\ \arctan(\frac{y}{x}) \end{bmatrix}, \quad \boldsymbol{C}(\boldsymbol{\theta}) = \begin{bmatrix} r^4 \sigma_r^2 & 0 \\ 0 & \sigma_\phi^2 \end{bmatrix} \quad (6.26)$$

where the term r^4 in the diagonal of range component of $\boldsymbol{C}(\boldsymbol{\theta})$ is used to model the effect that the amplitude of radar echo signal attenuates according to the fourth power of the target range; σ_r and σ_ϕ are the standard deviations of range and bearing measurement noise, respectively.

⁴Without confusion, subscripts for the temporal aspect of parameter $\boldsymbol{\theta}$ are dropped from the notation.

Using (6.24), we observe that the squared differential FID of Eq. (6.6) can be calculated as,

$$\begin{aligned}
ds^2 = & \sum_{i,j=1}^2 [\mathbf{G}(\boldsymbol{\theta})]_{ij} d\theta_i d\theta_j = \frac{1}{(x^2 + y^2)^2} \left[\frac{x^2}{(x^2 + y^2)\sigma_r^2} + \frac{y^2}{\sigma_\phi^2} + 8x^2 \right] dx^2 \\
& + \frac{1}{(x^2 + y^2)^2} \left[\frac{y^2}{(x^2 + y^2)\sigma_r^2} + \frac{x^2}{\sigma_\phi^2} + 8y^2 \right] dy^2 \\
& + \frac{2}{(x^2 + y^2)^2} \left[\frac{xy}{(x^2 + y^2)\sigma_r^2} - \frac{xy}{\sigma_\phi^2} + 8xy \right] dx dy
\end{aligned} \tag{6.27}$$

The IFID integral in (6.8) is obtained by calculating the geodesic that connects the locations of the two targets. This can be done by solving the Euler-Lagrange equations (6.14). In this example, the equations are of the following form:

$$\begin{aligned}
\frac{d^2x}{dt^2} + \Gamma_{11}^1 \left(\frac{dx}{dt}\right)^2 + 2\Gamma_{12}^1 \frac{dx}{dt} \frac{dy}{dt} + \Gamma_{22}^1 \left(\frac{dy}{dt}\right)^2 &= 0 \\
\frac{d^2y}{dt^2} + \Gamma_{11}^2 \left(\frac{dx}{dt}\right)^2 + 2\Gamma_{12}^2 \frac{dx}{dt} \frac{dy}{dt} + \Gamma_{22}^2 \left(\frac{dy}{dt}\right)^2 &= 0
\end{aligned} \tag{6.28}$$

where the Christoffel symbols of the second kind are (for $\sigma_r^2 = 1$, $\sigma_\phi^2 = 0.04$)

$$\begin{aligned}
\Gamma_{11}^1 &= -(8(x^2 + y^2)^2 + 2x^2 + y^2) x/a, \quad \Gamma_{11}^2 = (8(x^2 + y^2)^2 + y^2) y/a, \\
\Gamma_{12}^1 &= \Gamma_{21}^1 = \Gamma_{11}^1 \frac{y}{x}, \quad \Gamma_{12}^2 = \Gamma_{21}^2 = -(8(x^2 + y^2)^2 + x^2 + 2y^2) x/a, \\
\Gamma_{22}^1 &= (8(x^2 + y^2)^2 + x^2) x/a, \quad \Gamma_{22}^2 = \Gamma_{12}^2 \frac{y}{x}
\end{aligned} \tag{6.29}$$

with $a = (x^2 + y^2)^2(8x^2 + 8y^2 + 1)$.

The equations in (6.28) are second order nonlinear differential equations and can only be solved numerically. Each pair of initial conditions (a starting point $\boldsymbol{\theta}(0)$ and a tangent vector $\boldsymbol{\nu}$) corresponds to a unique solution of the geodesic. By setting initial conditions in the exponential map in (6.16), where $\boldsymbol{\theta}(0)$ corresponds a location and $\boldsymbol{\nu}(\varphi) = [\cos \varphi, \sin \varphi]^T$, $0 \leq \varphi \leq 2\pi$ is an unit tangent vector, the $\boldsymbol{\theta}(t)$ of a geodesic can be found by solving (6.28). A map of geodesics of identical lengths, that is, a circle in the IFID metric is sketched in Fig. 6.2. Such a map illustrates differences between the Euclidean distance and the information distance between two targets.

Fig. 6.3 (a) shows the map of geodesics for the sensor network of Example 1. Shown are the set of geodesics $\{\Psi(T; \boldsymbol{\theta}(0), \boldsymbol{\nu}(\varphi)), \varphi = 2\pi k/64, k = 0, \dots, 63\}$, with initial location $\boldsymbol{\theta}(0) = [10, 10]^T$, unit speed $|\boldsymbol{\nu}| = 1$ and end time $T = 20$. The end points form a manifold “circle” with radius $|\boldsymbol{\nu}|T$, which is the IFID along these geodesics. For a comparison, we also plot the “circles” formed by using the KLD as a measure of distance in Fig. 6.3 (b). The two “circles” are quite different in this case. It is important to note that, as illustrated in Fig. 6.3(d), the “equidistant” points measured by the IFID will generally not correspond to the same points measured in the target state (Euclidean) space. In Fig. 6.3(c), the “circle” of ED is plotted. This plot shows that: 1) *under the Riemannian metric, a free particle will generally travel non-equal distances along different geodesics to accumulate the same amount of energy*, and 2) the shape of the ED circle is in between that of IFID and KLD.

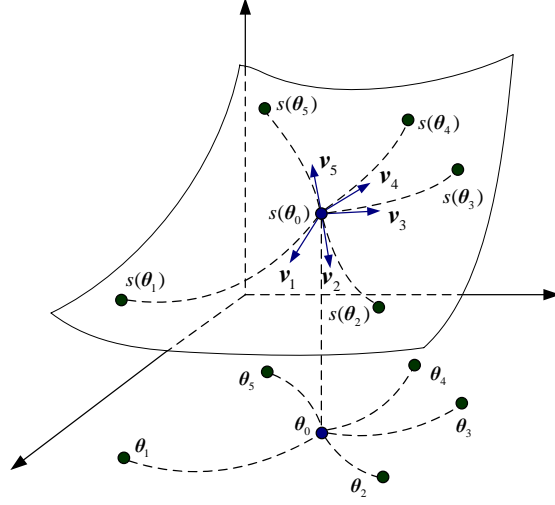


Figure 6.2: Illustration of geodesics that start from the same locations with initial unit tangent vectors.

Sensor network involving two bearings-only sensors – Example 2

In this example, we consider the problem of target localization using bearings-only measurements from two passive sensors. As shown in Fig. 6.4, the two passive sensors are located at (η_i, ξ_i) , $i = 1, 2$ and observe the bearings of a target subject to a Gaussian zero-mean random noise. The measurements satisfy (6.22) with the mean and covariance matrix given by

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{bmatrix} \arctan\left(\frac{y-\xi_1}{x-\eta_1}\right) \\ \arctan\left(\frac{y-\xi_2}{x-\eta_2}\right) \end{bmatrix}, \quad \mathbf{C}(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_{\phi_1}^2 & 0 \\ 0 & \sigma_{\phi_2}^2 \end{bmatrix} \quad (6.30)$$

where $\boldsymbol{\theta} = [x, y]^T$ is the target state vector and $\sigma_{\phi_1} = 0.2$ and $\sigma_{\phi_2} = 0.2$ are the standard deviations of the measurement noise for sensors 1 and 2, respectively.

The FIM in this case is derived as

$$g_{11}(\boldsymbol{\theta}) = \sum_{i=1}^2 \frac{1}{\sigma_{\phi_i}^2} \frac{(y - \xi_i)^2}{[(x - \eta_i)^2 + (y - \xi_i)^2]^2} \quad (6.31)$$

$$g_{12}(\boldsymbol{\theta}) = g_{21}(\boldsymbol{\theta}) = \sum_{i=1}^2 -\frac{1}{\sigma_{\phi_i}^2} \frac{(x - \eta_i)(y - \xi_i)}{[(x - \eta_i)^2 + (y - \xi_i)^2]^2} \quad (6.32)$$

$$g_{22}(\boldsymbol{\theta}) = \sum_{i=1}^2 \frac{1}{\sigma_{\phi_i}^2} \frac{(x - \eta_i)^2}{[(x - \eta_i)^2 + (y - \xi_i)^2]^2} \quad (6.33)$$

Fig. 6.5(a) shows the geodesics $\{\Psi(t; \boldsymbol{\theta}(0), \boldsymbol{\nu})\}$, where $t \in [0, 12]$, $\boldsymbol{\theta}(0) = [15, 15]^T$ and $\boldsymbol{\nu} = [\cos \varphi, \sin \varphi]^T$. The two passive sensors are located at $(\eta_1, \xi_1) = (0, 0)$ and $(\eta_2, \xi_2) = (50, 10)$. The IFIDs along all these geodesics are equal to

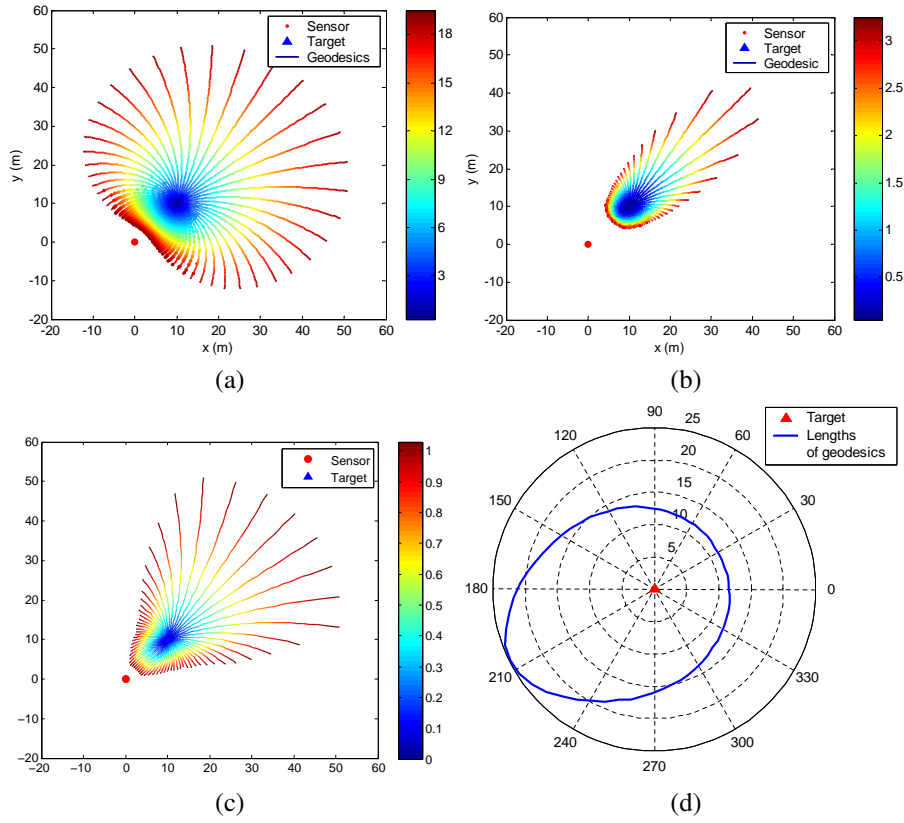


Figure 6.3: Figures of example 1: (a) FID circle; (b) KLD circle; (c) ED circle. All circles were drawn in the target state space Θ . (d) The lengths of IFID on the statistical manifold which corresponds to a circle centered at $\theta(0) = [10, 10]^T$ in the target state (Euclidean) space.

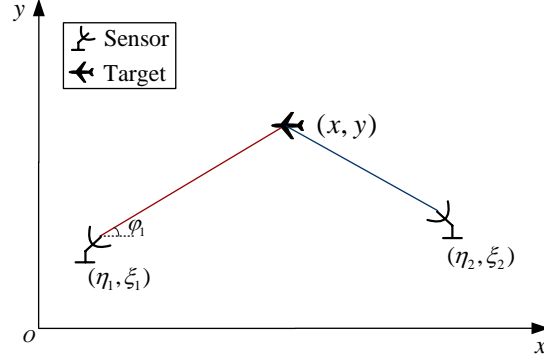


Figure 6.4: The measurement model of bearings-only measurements of two sensors.

$|\nu|T = 12$. The set of end points of geodesics describes a manifold circle of identical FIDs centered at the target location $\theta(0)$. For comparison, the KLD circle that centered at $\theta(0)$ is also drawn in Fig. 6.5(c). The plots are repeated at a different initial target state $\theta(0) = [40, 25]^T$ in Figs. 6.5(b) and 6.5(d). These plots highlight the different responses of the IFID and KLD to changes in the target location in this sensor network scenario. The ED circles for the above two cases are plotted in Fig. 6.5(e) and 6.5(f).

Sensor network involving three ranges-only sensors – Example 3

In this example, we consider an extended example of the target localization problem in a sensor network of three range-only sensors. As shown in Fig. 6.6, these three range-only sensors are located at (η_i, ξ_i) , $i = 1, 2, 3$ and observe the ranges between a target and the sensors subject to a random range noise. In this network configuration, the likelihood function is described by (6.22) with

$$\mu(\theta) = \begin{bmatrix} \sqrt{(x - \eta_1)^2 + (y - \xi_1)^2} \\ \sqrt{(x - \eta_2)^2 + (y - \xi_2)^2} \\ \sqrt{(x - \eta_3)^2 + (y - \xi_3)^2} \end{bmatrix}, \quad C(\theta) = \begin{bmatrix} r_1^4 \sigma_{r_1}^2 & 0 & 0 \\ 0 & r_2^4 \sigma_{r_2}^2 & 0 \\ 0 & 0 & r_3^4 \sigma_{r_3}^2 \end{bmatrix} \quad (6.34)$$

The FIM for the measurement model (6.34) can be derived as

$$g_{11}(\theta) = \sum_{i=1}^3 \frac{(x - \eta_i)^2}{r_i^4} \left[\frac{1}{r_i^2 \sigma_{r_i}^2} + 8 \right] \quad (6.35)$$

$$g_{12}(\theta) = g_{21}(\theta) = \sum_{i=1}^3 \frac{(x - \eta_i)(y - \xi_i)}{r_i^4} \left[\frac{1}{r_i^2 \sigma_{r_i}^2} + 8 \right] \quad (6.36)$$

$$g_{22}(\theta) = \sum_{i=1}^3 \frac{(y - \xi_i)^2}{r_i^4} \left[\frac{1}{r_i^2 \sigma_{r_i}^2} + 8 \right] \quad (6.37)$$

As in previous examples, the manifold circles of identical IFIDs and KLDs are plotted in Fig. 6.7(a) and 6.7(c), where the geodesics start from at $\theta(0) = [20, 10]^T$ and

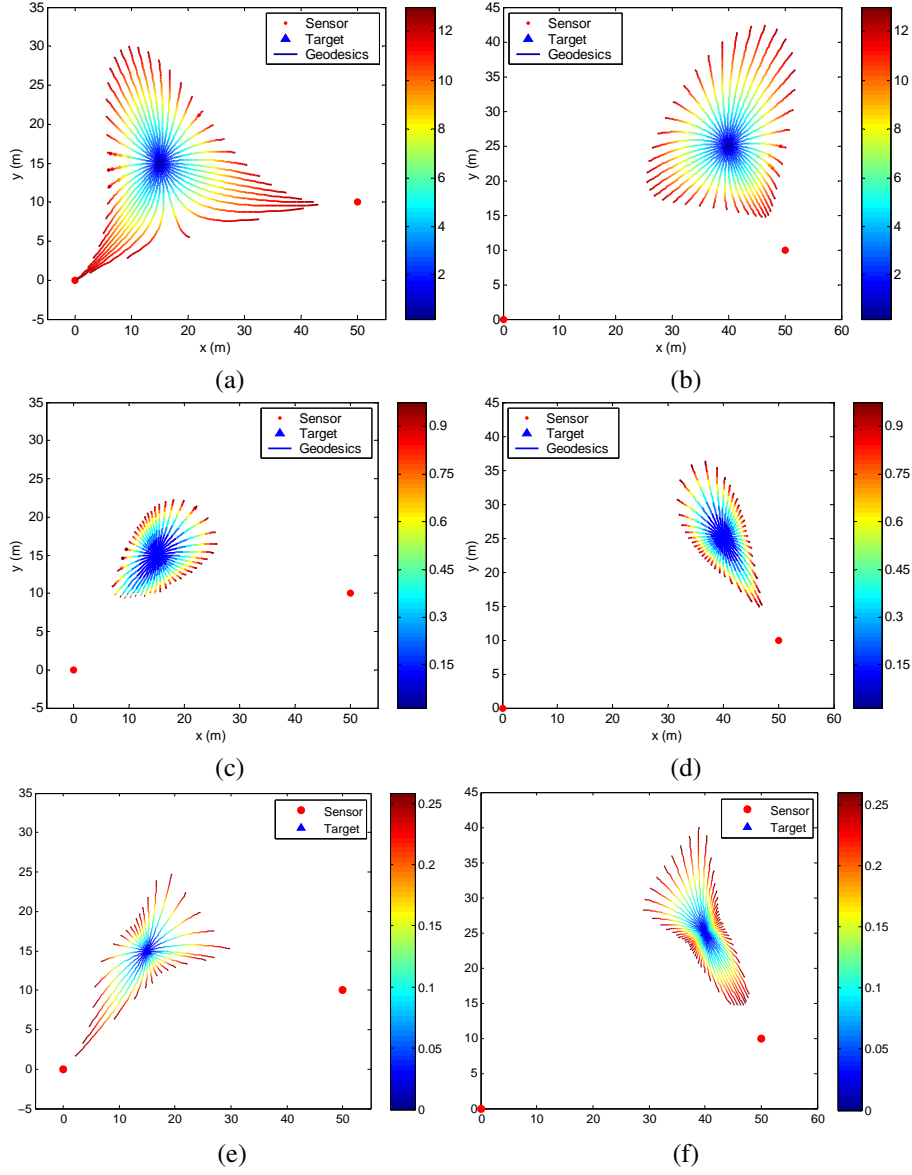


Figure 6.5: Figures of Example 2: (a), (c) and (e) are respectively the IFID, KLD and ED circles at $\theta(0) = [15, 15]^T$ drawn in Θ where the two passive sensors are located at $(\eta_1, \xi_1) = (0, 0)$ and $(\eta_2, \xi_2) = (50, 10)$, respectively. (b), (d) and (f) are the replicated plots of (a), (c) and (e) respectively at $\theta(0) = [40, 25]^T$.

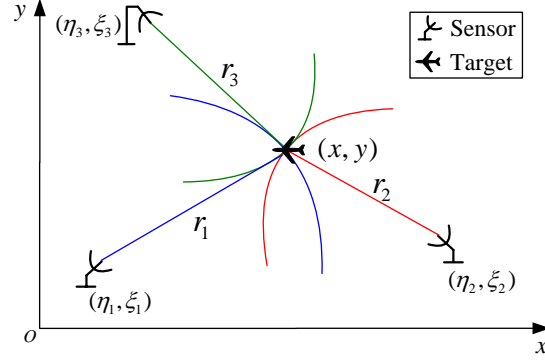


Figure 6.6: Example of sensor network of three range only sensors for target localisation

end with identical lengths of $|\nu|T = 15$. The plot is repeated in Fig. 6.7(b) and 6.7(d), where all geodesics start from $\theta(0) = [-10, 10]^T$ with identical lengths of $|\nu|T = 10$. In this example, the three sensors are located at $(\eta_1, \xi_1) = (0, 0)$, $(\eta_2, \xi_2) = (15, 30)$ and $(\eta_3, \xi_3) = (50, 10)$ with the standard deviations of noise $\sigma_{r_1} = 1$, $\sigma_{r_2} = 2$ and $\sigma_{r_3} = 3$, respectively. The ED circles centered at the aforementioned two locations are given in Fig. 6.7(e) and 6.7(f) respectively.

It is interesting to verify the behavior of the geodesics in a large scale in this example. Fig. 6.8 shows two geodesics $\{\Psi(t; \theta(0), \nu_i)\}$, $i = 1, 2$, $0 \leq t \leq 150$, which start from the same location $\theta(0) = [-10, 10]^T$ and have initial tangent vectors $\nu_i = [\cos \varphi_i, \sin \varphi_i]^T$ with $\varphi_1 = 0$ and $\varphi_2 = \pi/2$. Clearly, as illustrated in Fig. 6.8, the geodesic between two points A and B in a Riemannian manifold is not unique. In this, and every case, the IFID between two points corresponds to the geodesic of shortest length.

Remarks:

1. As demonstrated in Figs. 6.3, 6.5, and 6.7, an IFID “circle” of a statistical manifold generally does not correspond to a circle in Euclidean (target state) space and vice versa.
2. Geodesics between two points in a statistical manifold are not unique; the IFID between two points is the length of the shortest geodesic.
3. One of the main differences between the IFID and the KLD is that the IFID measures the distance (i.e., the length of the shortest geodesic) between two points on a statistical manifold and it is a genuine metric while the KLD is not.
4. IFID may be used to measure the underlying sensor ability to resolve closely spaced targets. In practice, a threshold of the minimum IFID required for separating two closely spaced targets may be set. Two closely spaced targets cannot be resolved from a measurement if the IFID between them is below the threshold. This concept is illustrated in Fig. 6.9 in terms of a “resolution cell” in the sensor network of Example 3. All edges of the colored areas represent identical

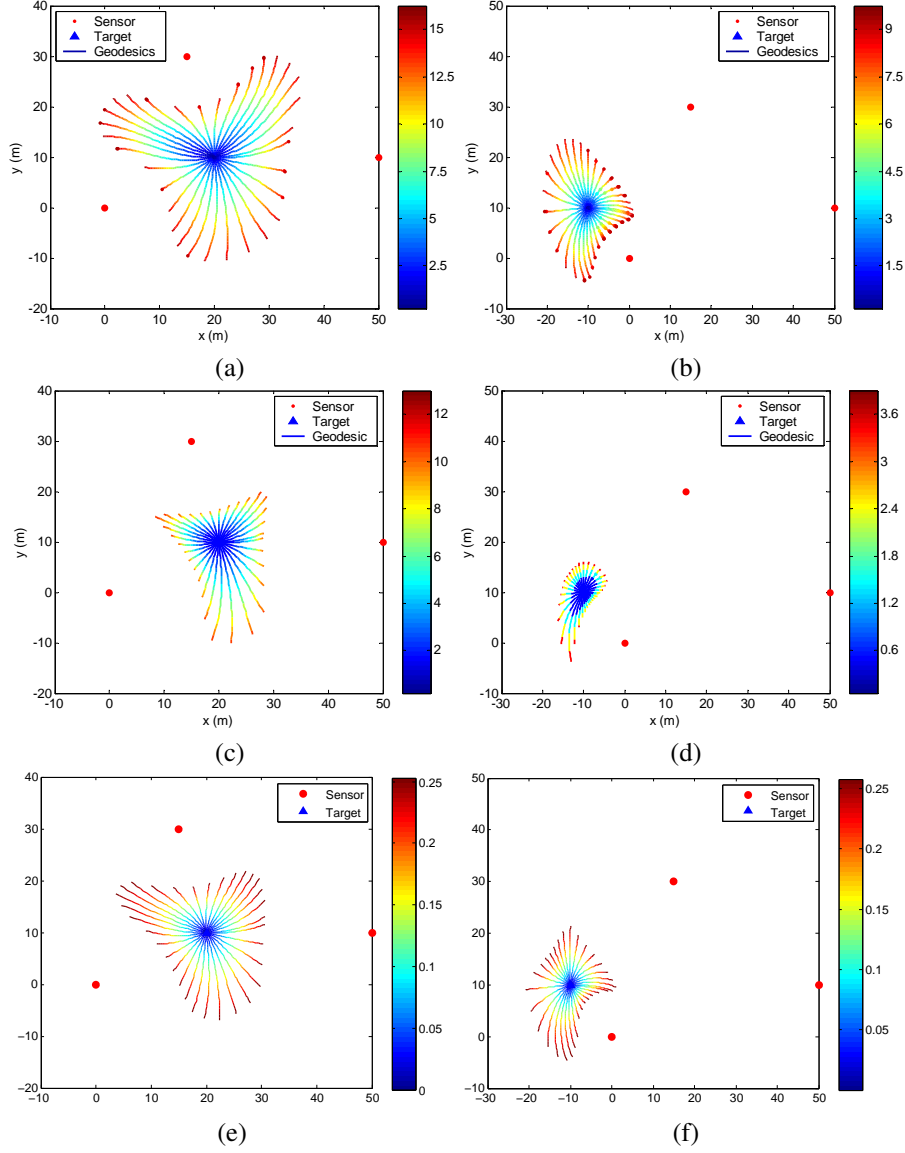


Figure 6.7: Figures of Example 3: (a), (c) and (e) are the circles of FID, KLD and ED respectively drawn in Θ at $\theta(0) = [20, 10]^T$. (b), (d) and (f) are the replicated plots of (a), (c) and (e) respectively at another location $\theta(0) = [-10, 10]^T$. Three sensors are located at $(\eta_1, \xi_1) = (0, 0)$, $(\eta_2, \xi_2) = (15, 30)$ and $(\eta_3, \xi_3) = (50, 10)$.

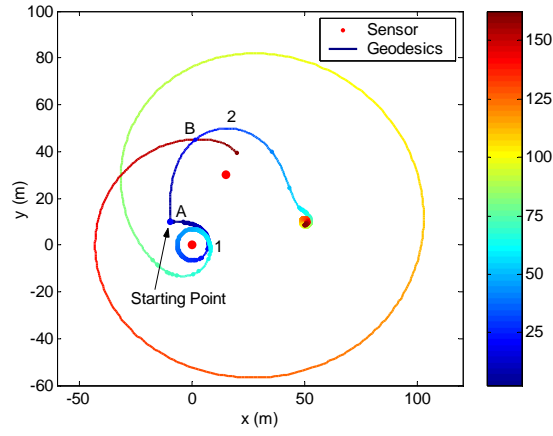


Figure 6.8: Illustration of two geodesics in Example 3 which start from the same location in different directions and are of the same length.

IFIDs between two the targets with one at the center and another at the edge of a colored area.

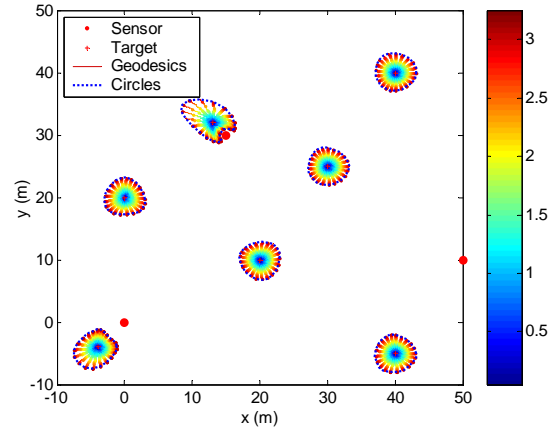


Figure 6.9: Illustration of FID circles which may serve as the “resolution cells” for the underlying sensor network to measure its ability to distinguish two closely spaced targets.

6.4.2 Riemannian and scalar curvatures

Figure 6.10 depicts the Ricci scalar curvatures of the statistical manifold in the sensor network of three range-only sensors (i.e., Example 3). In this example, the Ricci scalar curvatures of the statistical manifold are less than or equal to zero. The plot provides an additional graphical view of properties of the underlying sensor network. For example, it reflects the rate of change of information which can be collected by the sensor network at a particular point.

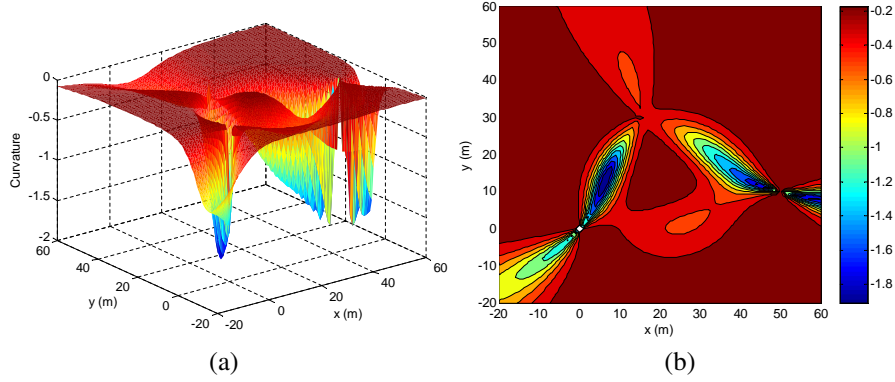


Figure 6.10: Ricci scalar curvatures of the statistical manifold for the sensor network of Example 3. (a) Ricci scalar curvature distribution. (b) Colour map of the scalar curvature.

6.4.3 Ricci curvature tensor field

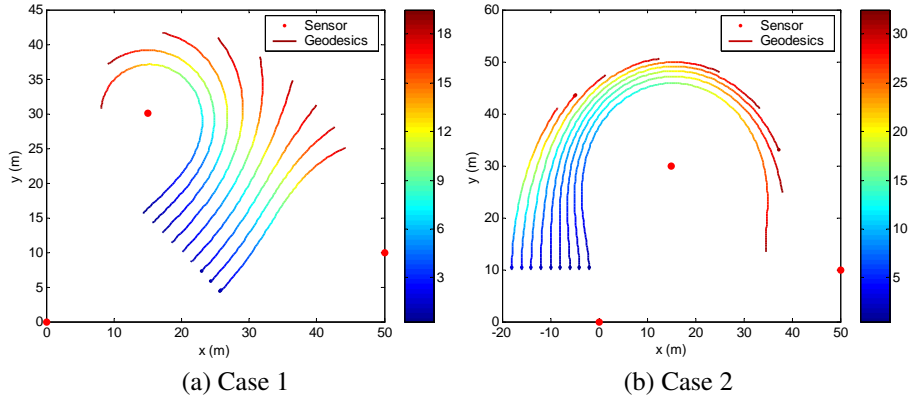


Figure 6.11: Bundle of geodesics of identical FIDs with parallel initial tangent vectors for the sensor network of three ranges-only sensors. Case 1 illustrates the situation of divergent geodesic bundle and a convergent bundle example is shown in Case 2.

Figure 6.11 shows a bundle of geodesics of identical FIDs with parallel initial tangent vectors in the sensor network of Example 3. Clearly, the bundle of geodesics in Case 1 deviate from each other in the target state plane whereas the bundle of geodesics in Case 2 are of another type of deviation, i.e., they deviate from each other in a direction that is perpendicular to the plot plane in some region. This behavior of a bundle of initially parallel geodesics indicates the structure of the manifold corresponding to an underlying sensor network. It is related to the information change rate of the sensor network, which can be measured by the Ricci curvature.

Figure 6.12 represents the Ricci curvature tensor field of the statistical manifold of Example 3. Taking sign into account, the Ricci curvature tensor field can be regarded as information ellipses which indicate both the amplitude and direction of the information change rate observable to the network. This information is important for target tracking

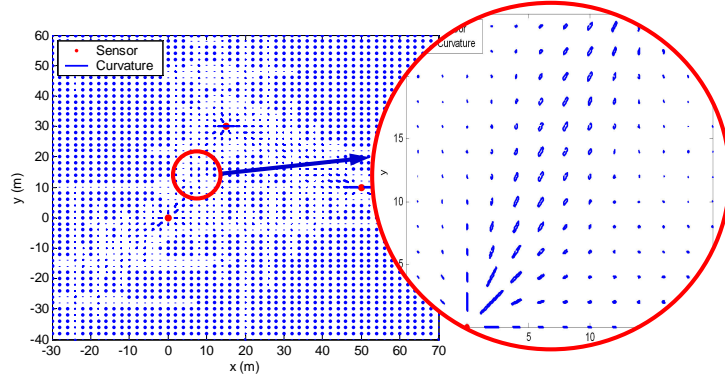


Figure 6.12: Ricci curvature tensor field of the manifold with range-only measurements from three sensors.

in the sensor network. In Figure 6.13, the surface of the determinant of the FIM, which serves as the metric in the statistical manifold of Example 3, is plotted on a logarithm scale. It illustrates the amount of information that can be acquired by the sensor network when the target is in a particular position.

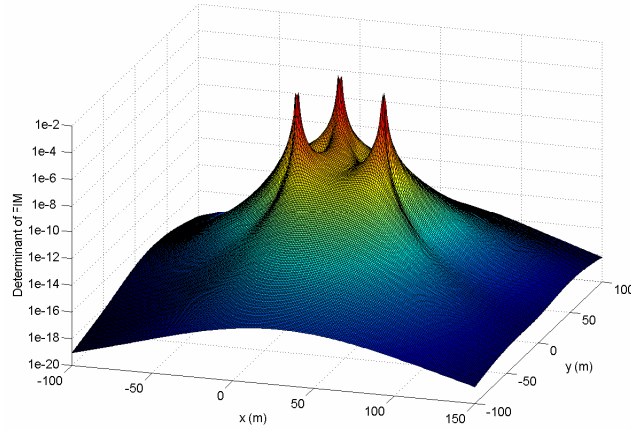


Figure 6.13: The surface of the determinant of the FIM in Example 3 plotted on a logarithm scale.

Discussions:

1. As demonstrated in Figure 6.10, the Ricci scalar curvatures of the statistical manifold in Example 3 are negative. Interestingly, the scalar curvatures of the statistical manifold of the other two examples vanish everywhere. A manifold of zero Ricci curvature is called a Ricci-flat manifold and it indicates the manifold is locally flat everywhere, i.e., the geodesic ball of the manifold is geometrically identical to the standard ball in Euclidean space locally. This can also be seen from Equation (6.21), where the vanishing of Ricci curvature for a sensor

network reflects the fact that the Fisher information volume near a point on the statistical manifold has the same geometric shape as that described in Euclidean space.

2. Examples 2 and 3 may be generalised to a N -sensor case (i.e., $N = 2, 3, \dots$ for Example 2 and $N = 3, 4, \dots$ for Example 3). Our calculations indicate that the Ricci scalar curvatures can be positive when $N > 3$ for Example 3. The statistical manifold is no longer Ricci flat in the case of $N > 2$ sensors as presented in Example 2. These observations suggest a strategic optimisation method which is based on the required values of Ricci scalar curvature for the sensor placement problem in sensor network design. This topic is currently being investigated as continuing research work.
3. As demonstrated in Fig. 6.12, the Ricci curvature tensor provides the information change rate along a given direction on the manifold with zero Ricci curvature reflecting an isotropic change rate. In addition, it indicates the amount of information able to be collected by the underlying sensor network.
4. In many sensor network optimisation problems, such as sensor scheduling, radar waveform design etc., the amount of information collected by the network sensors serves as an important criterion. From the information geometry point of view, the optimisation is achieved by changing the associated statistical manifold structure, which is described by Riemannian curvatures. In [67] the potential of optimal sensor scheduling via the information geometry has been demonstrated.

6.5 Statistical Manifold Representation

The shape of statistical manifolds of higher dimensions is difficult to describe in terms of spaces that we are familiar with. A reasonable way is to seek a representation of the manifold as an immersion in the Euclidean space \mathbb{R}^n , which will preserve the differential structure of the original manifold and will have a derivative which is everywhere injective. The affine immersion discussed by Dodson and Matsuzoe in [87] provides such a framework in which the underlying statistical manifold of sensor networks can be realised in \mathbb{R}^n . In this section, we will discuss the general affine immersion for the multivariate Gaussian manifold corresponding to our sensor network examples.

6.5.1 Exponential family of probability density functions

As we mentioned earlier, the measurement of the sensors we discussed obeys a multivariate normal distribution which belongs to the exponential family of distributions. The exponential family, which includes popular distributions such as the Gaussian, Poisson and Gamma, is widely used in probability and statistics due to its many desirable properties [69].

An n -dimensional set of probability density functions $S = \{p_{\theta} | \theta \in \Theta \subset \mathbb{R}^n\}$ is said to be an exponential family when the density functions can be expressed in terms

of functions $\{C, F_1, \dots, F_n\}$ and a function φ on Θ such that [88]:

$$p_{\theta}(\mathbf{x}) = \exp\left\{C(\mathbf{x}) + \sum_i (\theta_i F_i(\mathbf{x})) - \varphi(\theta)\right\} \quad (6.38)$$

where $\mathbf{x} \in \Omega$ is a vector valued measurement, $\theta = [\theta_1, \dots, \theta_n]^T$ are the natural coordinates or canonical parameters, $C(\mathbf{x})$ represents the exponential component which is independent of θ and $\mathbf{F}(\mathbf{x}) = [F_1(\mathbf{x}), \dots, F_n(\mathbf{x})]^T$ are sufficient statistics for θ . The function φ is called the potential function of the exponential family and it is found from the normalisation condition $\int_{\Omega} p_{\theta}(\mathbf{x}) d\mathbf{x} = 1$, i.e.,

$$\varphi(\theta) = \log \int_{\Omega} \exp\left\{C(\mathbf{x}) + \sum_i (\theta_i F_i(\mathbf{x}))\right\} d\mathbf{x} \quad (6.39)$$

From the definition of an exponential family, and with $\partial_i = \partial/\partial\theta_i$, we use the log-likelihood function $l(\theta, \mathbf{x}) = \log(p_{\theta}(\mathbf{x}))$ to obtain

$$\partial_i l(\theta, \mathbf{x}) = F_i(\mathbf{x}) - \partial_i \varphi(\theta) \quad (6.40)$$

and

$$\partial_i \partial_j l(\theta, \mathbf{x}) = -\partial_i \partial_j \varphi(\theta) \quad (6.41)$$

The Fisher information metric \mathbf{G} on the n -dimensional space of parameters $\Theta \subset \mathbb{R}^n$, equivalently on the set $S = \{p_{\theta} | \theta \in \Theta \subset \mathbb{R}^n\}$, has coordinates [89]:

$$g_{ij} = - \int_{\Omega} \partial_i \partial_j l(\theta, \mathbf{x}) p_{\theta}(\mathbf{x}) d\mathbf{x} = \partial_i \partial_j \varphi(\theta) \quad (6.42)$$

Then, (S, g) is a Riemannian n -manifold with Levi-Civita connection given by [89]:

$$\begin{aligned} \Gamma_{ij}^k(\theta) &= \sum_{l=1}^n \frac{1}{2} g^{kl} (\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}) \\ &= \sum_{l=1}^n \frac{1}{2} g^{kl} \partial_i \partial_j \partial_l \varphi(\theta). \end{aligned} \quad (6.43)$$

where $[g^{kl}]$ represents the inverse of $[g_{kl}]$.

The justification of the underlying statistical manifold representation by immersion (in the natural parameter θ) in an exponential family can be viewed from the following two points:

1. Any exponential family of distributions has a unique a potential function and the latter completely describes this exponential family of distributions [90]. In other words, we can fully understand the mean and covariance of the sufficient statistics $F_i(\mathbf{x})$, $i = 1, \dots, n$ in (6.38) by differentiating $\varphi(\theta)$, i.e.,

$$E\{F_i(\mathbf{x})\} \triangleq \boldsymbol{\eta} = \partial_i \varphi(\theta) \quad (6.44)$$

$$E\{[F_i(\mathbf{x}) - E(F_i(\mathbf{x}))][F_j(\mathbf{x}) - E(F_j(\mathbf{x}))]^T\} = \partial_i \partial_j \varphi(\theta) \quad (6.45)$$

2. The Fisher information matrix (6.42) and thus the connection (Christoffel symbols) of (6.43) corresponding to the statistical manifold can all be described using the potential function $\varphi(\theta)$.

6.5.2 Affine immersion of a manifold

Let M be an n -dimensional manifold, and f an immersion from M to \mathbb{R}^{n+1} , i.e., the differential map

$$D_p f : \mathcal{T}_p M \implies \mathcal{T}_{f(p)} \mathbb{R}^{n+1} \quad (6.46)$$

is an injective map in the tangent space, denoted as \mathcal{T}_p , at every point $p \in M$. Suppose that ξ is a local vector field along f . The pair $\{f, \xi\}$ is said to be an affine immersion from M to \mathbb{R}^{n+1} if, for each point $p \in M$, the following formula holds [90]:

$$\mathcal{T}_{f(p)} \mathbb{R}^{n+1} = f_*(\mathcal{T}_p M) \bigoplus \text{Span}\{\xi_p\} \quad (6.47)$$

where $\mathcal{T}_{f(p)} \mathbb{R}^{n+1}$ can be identified with \mathbb{R}^{n+1} , $\forall f(p) \in \mathbb{R}^{n+1}$, \bigoplus denotes the direct sum of two subspaces and Span is the span operator on a collection of vectors in linear algebra. We call ξ a transversal vector field. Eq. (6.47) is a technical requirement to ensure that the differential structure is preserved into the immersion.

For manifolds of multivariate Gaussian distributions, which is the case of our sensor network scenarios, the representation of manifold can be realized in Euclidean space \mathbb{R}^{n+1} by the following affine immersion:

Proposition 6.5.1. *Let M be the multivariate Gaussian manifold with the Fisher information metric g . Denote by $(\boldsymbol{\theta}, \boldsymbol{\Xi})$ a natural coordinate system. Then M can be realized in \mathbb{R}^{n+1} by the graph of a potential function, namely, M can be realized by the affine immersion $\{f, \xi\}$:*

$$f : M \rightarrow \mathbb{R}^{n+1} : \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\Xi} \end{bmatrix} \mapsto \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\Xi} \\ \varphi \end{bmatrix}, \quad \xi = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (6.48)$$

where φ is the potential function.

We give a simple 1D example in the next sub-section to demonstrate how to find the potential function from a given distribution.

6.5.3 An example of a gaussian 2-manifold

The family of univariate normal or Gaussian density functions has an event space $\Omega = \mathbb{R}$ and the probability density functions are given by

$$M = \left\{ p(x; \mu, \sigma^2) | p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+ \right\} \quad (6.49)$$

The mean μ and standard deviation σ are frequently used as a local coordinate system $(\xi_1, \xi_2) = (\mu, \sigma)$ as in [90].

The univariate Gaussian density can be written as

$$\begin{aligned}
p(x; \theta_1, \theta_2) &= \exp \left\{ \frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 - \frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2) \right\} \\
&= \exp \left\{ \theta_1 x + \theta_2 x^2 - \left[-\frac{\theta_1^2}{4\theta_2} + \frac{1}{2} \log\left(-\frac{\pi}{\theta_2}\right) \right] \right\} \\
&= \exp \{ C(x) + F_1(x)\theta_1 + F_2(x)\theta_2 - \varphi(\theta_1, \theta_2) \}
\end{aligned} \tag{6.50}$$

In the Gaussian 2-manifold, set $\theta_1 = \frac{\mu}{\sigma^2}$ and $\theta_2 = -\frac{1}{2\sigma^2}$. Then $(\theta_1, \theta_2) = (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$ is the natural coordinate system and

$$\varphi = -\frac{\theta_1^2}{4\theta_2} + \frac{1}{2} \log\left(-\frac{\pi}{\theta_2}\right) = \frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi}\sigma) \tag{6.51}$$

is the corresponding potential function.

For a natural coordinate system, there exists a function φ (potential function) on M such that the Fisher information metric is given by the Hessian of φ , that is

$$\frac{\partial^2 \varphi}{\partial \theta_i \partial \theta_j} = g_{ij} \tag{6.52}$$

where $[g_{ij}]$ is the Fisher information metric with respect to the natural coordinate system.

According to (6.52), the Fisher metric of the Gaussian 2-manifold with respect to natural coordinates (θ_1, θ_2) is given by

$$[g_{ij}] = \begin{bmatrix} \frac{-1}{2\theta_2} & \frac{\theta_1}{2\theta_2^2} \\ \frac{\theta_1}{2\theta_2^2} & \frac{\theta_2 - \theta_1^2}{2\theta_2^3} \end{bmatrix} = \begin{bmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^2(2\mu^2 + \sigma^2) \end{bmatrix} \tag{6.53}$$

6.5.4 Manifold representations of sensor networks

For the sensor network examples discussed in this report, the measurement errors are characterized with multivariate Gaussian distributions, i.e.,

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\xi}), \mathbf{C}(\boldsymbol{\xi})) \tag{6.54}$$

where $\boldsymbol{\xi} = [x, y]^T \in \boldsymbol{\Theta}$ is the state of interest in the local coordinate system.

The probability density function of the measurement error is

$$p(\mathbf{x}; \boldsymbol{\xi}) = |2\pi\mathbf{C}(\boldsymbol{\xi})|^{-1/2} \exp\{-[\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\xi})]^T \mathbf{C}^{-1}(\boldsymbol{\xi}) [\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\xi})]/2\} \tag{6.55}$$

It can be represented in exponential form using the collection of sufficient statistics $(\mathbf{x}, \mathbf{x}\mathbf{x}^T)$. Let $\boldsymbol{\theta}$ be a m -vector of parameters associated with the vector of sufficient statistics $\mathbf{x} = [x_1, \dots, x_m]^T$, where m is the dimension of \mathbf{x} , and a symmetric matrix $\boldsymbol{\Xi} \in \mathbb{R}^{m \times m}$ associated with the matrix $\mathbf{x}\mathbf{x}^T$. Then the multivariate Gaussian is an exponential family of the form [91]:

$$p(\mathbf{x}; \boldsymbol{\theta}) = \exp \{ \langle \boldsymbol{\theta}, \mathbf{x} \rangle + \langle \boldsymbol{\Xi}, \mathbf{x}\mathbf{x}^T \rangle - \varphi(\boldsymbol{\theta}, \boldsymbol{\Xi}) \} \tag{6.56}$$

where $\langle \boldsymbol{\theta}, \mathbf{x} \rangle = \sum_{i=1}^m \theta_i x_i$ is the Euclidean inner product on \mathbb{R}^m , and

$$\langle \langle \boldsymbol{\Xi}, \mathbf{x}\mathbf{x}^T \rangle \rangle = \text{tr}(\boldsymbol{\Xi}\mathbf{x}\mathbf{x}^T) = \sum_{i=1}^m \sum_{j=1}^m \Xi_{ij} x_i x_j \quad (6.57)$$

is the Frobenius inner product on symmetric matrices.

We may write the natural parameters of mixed-type $(\boldsymbol{\theta}, \boldsymbol{\Xi}) = (\mathbf{C}^{-1}\boldsymbol{\mu}, -\frac{1}{2}\mathbf{C}^{-1})$ with the corresponding potential function [92]

$$\varphi(\boldsymbol{\theta}, \boldsymbol{\Xi}) = -\frac{1}{4}\text{tr}(\boldsymbol{\Xi}^{-1}\boldsymbol{\theta}\boldsymbol{\theta}^T) - \frac{1}{2}\log |-\boldsymbol{\Xi}| + \frac{m}{2}\log \pi \quad (6.58)$$

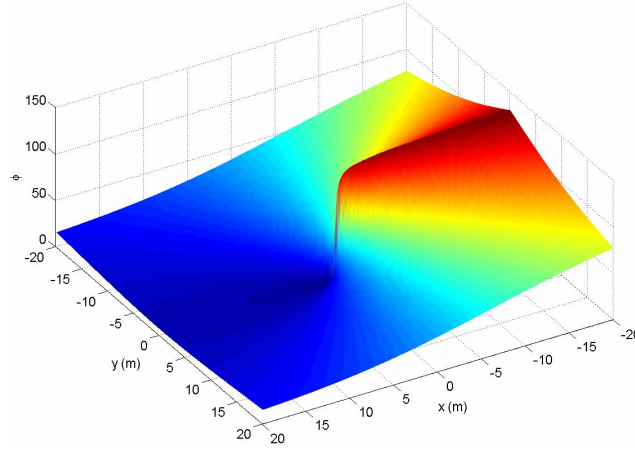


Figure 6.14: Affine immersion for the manifold of single conventional radar network in Example 1.

The potential function $\varphi(\boldsymbol{\theta}, \boldsymbol{\Xi})$ is a strictly convex and differentiable function that specifies uniquely the exponential family. The one-to-one mapping from the original parameters $(\boldsymbol{\mu}, \mathbf{C})$ to natural parameters $(\boldsymbol{\theta}, \boldsymbol{\Xi})$ is given by

$$\begin{bmatrix} \boldsymbol{\mu} \\ \mathbf{C} \end{bmatrix} \Leftrightarrow \begin{bmatrix} \mathbf{C}^{-1}\boldsymbol{\mu} \\ -\frac{1}{2}\mathbf{C}^{-1} \end{bmatrix} \quad (6.59)$$

Therefore, the potential function can be expressed in terms of local parameters as follows:

$$\varphi(\boldsymbol{\theta}, \boldsymbol{\Xi}) = \frac{1}{2}\boldsymbol{\mu}^T \mathbf{C}^{-1} \boldsymbol{\mu} + \frac{1}{2}\log |\mathbf{C}| + \frac{m}{2}\log 2\pi. \quad (6.60)$$

Figures 6.14–6.16 show affine immersions in \mathbb{R}^3 for the manifolds of the three basic forms of sensor network examples discussed previously. Intuitively, these plots indicate the maximum relative error of the parameters in the applications to sensor networks.

Discussions:

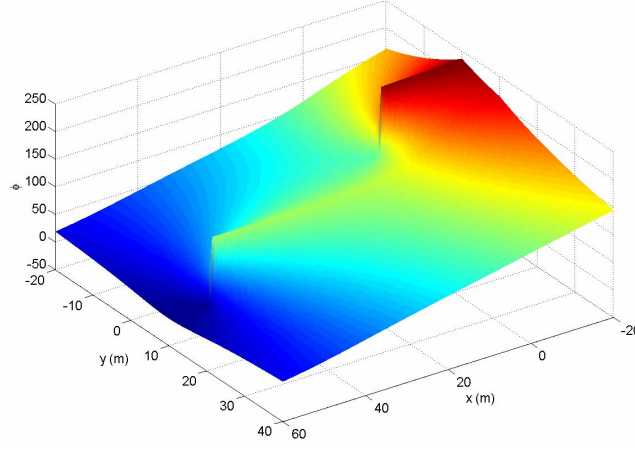


Figure 6.15: Affine immersion for the manifold of two bearings-only sensor network in Example 2.

1. For distributions of the exponential family, a statistical manifold can be represented by its potential function in natural parameters via an immersion which preserves the differential structure of the statistical manifold. In our examples as illustrated in Figures 6.14, 6.15 and 6.16, the manifold representation, given by the potential function $\varphi(\boldsymbol{\theta}, \boldsymbol{\Xi})$ in Eq. (6.60), forms a curved surface in terms of local parameter $[x, y]^T$. Their values – under the re-parameterized parameters $(\boldsymbol{\theta}, \boldsymbol{\Xi})$, indeed, signify *possible relative error* (or uncertainty) of the underlying sensor network measurement, that is, the ratio of measurement variance to the measurement value.
2. The lower the surface $\varphi(\boldsymbol{\theta}, \boldsymbol{\Xi})$ is, the smaller the possible relative error of a measurement is. In particular, as shown in Fig. 6.14 and Fig. 6.15, since the measurement model involves bearings, the value of φ will be higher when the bearing measurement value is smaller.
3. We observed that the manifold representations of all three examples have negative scalar Ricci curvatures everywhere. Generally speaking, the manifold of a Gaussian distribution in natural parameters results in a negative constant curvature, while the manifold of a curved Gaussian, as in the examples presented in this report, is of negative curvature everywhere. The curvatures of statistical manifolds are closely related to the state estimation problem [84] and we will address this issue for sensor networks in separate research work.

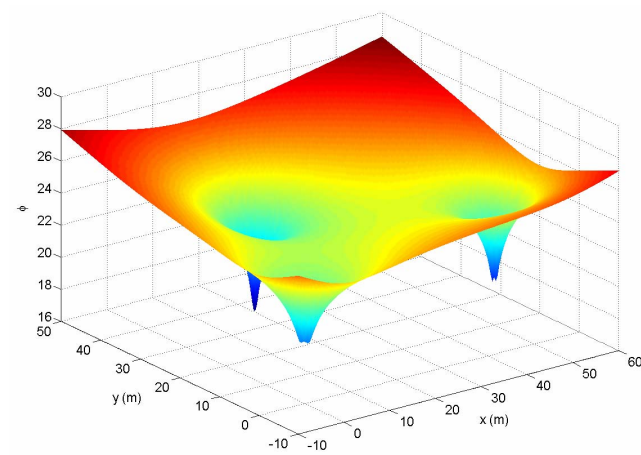


Figure 6.16: Affine immersion for the manifold of three ranges-only sensor network in Example 3.

Chapter 7

Concluding Remarks

We have studied the detection performance for trees with unbounded height. For balanced binary relay trees, the total error probability decays to 0 in the rate of \sqrt{N} , even if the sensors are asymptotically crummy. In addition, the scaling law for the decay rate remains in the case where sensors fail with certain probabilities. In the case where the communication links fail with certain probabilities, if all the links fail with identical probability, then the decay rate is strictly smaller than the non-failure case. If the link failure probabilities decay to 0 sufficiently fast towards the fusion center, then the scaling law for the decay rate remains. We further investigate the overall strategy which achieves the maximum of the reduction in the total error probability. We provide the explicit solution using dynamic programming method and Bellman's principle. We show that the reduction in the total error probability is a submodular function. Therefore, the greedy strategy achieves a total error probability that is at least a factor of the error probability achieved by the overall optimal strategy. We further study the detection performance of M -ary relay trees, which is a more general architecture. The impact of non-binary message alphabet is also investigated.

In this report, the use of information geometry theory in the performance evaluation of sensor networks is considered and the potential application of information geometry in sensor network analysis and design is demonstrated using three basic types of sensor network scenarios. In addition, an Euclidean immersion method for the representation of a statistical manifold is presented. The analysis results obtained suggest that geometrical constructs of statistical manifolds such as geodesics, Fisher information distance and curvature can be useful for the evaluation of the sensor measurement process and may facilitate sensor network design, evaluation and optimisation. The results presented in this report also highlight that information geometry offers a consistent and comprehensive means for understanding and solving sensor network issues in target tracking.

One limitation of our results is that it applies to particular architecture. Future research includes analyzing more general architectures. For example, un-balanced hierarchical architecture, trees with non-uniform degree, etc. For the communication link failure case, our model is essentially a deletion channel model for balanced binary relay trees. We can further consider symmetric channel, noisy channel, even a stucked

on/off channel for balanced binary relay trees and more general architectures. Another challenging question is the correlated sensor measurement case.

Appendix A

Proof of Proposition 2.3.1

If $(\alpha_k, \beta_k) \in B_m$, where m is a positive integer and $m \neq 1$, then

$$\frac{L_{k+1}}{L_k^2} = \frac{1 - (1 - \alpha_k)^2 + \beta_k^2}{(\alpha_k + \beta_k)^2}.$$

The following calculation establishes the lower bound of the ratio L_{k+1}/L_k^2 :

$$\begin{aligned} L_{k+1} - L_k^2 &= 1 - (1 - \alpha_k)^2 + \beta_k^2 - (\alpha_k + \beta_k)^2 \\ &= -2\alpha_k^2 - 2\alpha_k\beta_k + 2\alpha_k \\ &= 2\alpha_k(1 - (\alpha_k + \beta_k)) \geq 0, \end{aligned}$$

which holds in B_m .

To show the upper bound of the ratio L_{k+1}/L_k^2 , it suffices to prove that

$$\begin{aligned} L_{k+1} - 2L_k^2 &= 1 - (1 - \alpha_k)^2 + \beta_k^2 - 2(\alpha_k + \beta_k)^2 \\ &= -3\alpha_k^2 - 4\alpha_k\beta_k + 2\alpha_k - \beta_k^2 \leq 0. \end{aligned}$$

The partial derivative with respect to β_k is

$$\frac{\partial(L_{k+1} - 2L_k^2)}{\partial\beta_k} = -2\beta_k - 4\alpha_k \leq 0,$$

which is non-positive, and so it suffices to consider values on the upper boundary of B_1 .

$$\begin{aligned} L_{k+1} - 2L_k^2 &= 1 - (1 - \alpha_k)^2 + \beta_k^2 - 2(\alpha_k + \beta_k)^2 \\ &= 2\beta_k^2 - 2(\alpha_k + \beta_k)^2 \leq 0. \end{aligned}$$

In consequence, the claimed upper bound on the ratio L_{k+1}/L_k^2 holds.

Appendix B

Proof of Proposition 2.3.2

From Proposition 2.3.1 we have, for $k = 0, 1, \dots, m-2$,

$$L_{k+1} = a_k L_k^2$$

for some $a_k \in [1, 2]$. Then for $k = 1, 2, \dots, m-1$,

$$L_k = a_{k-1} \cdot a_{k-2}^2 \dots a_0^{2^{k-1}} L_0^{2^k},$$

where $a_i \in [1, 2]$ for each i . Hence,

$$\begin{aligned} \log L_k^{-1} &= -\log a_{k-1} - 2 \log a_{k-2} - \dots \\ &\quad - 2^{k-1} \log a_0 - \log L_0^{2^k}. \end{aligned}$$

Since $\log L_0^{-1} > 0$ and $0 \leq \log a_i \leq 1$ for each i , we have

$$\log L_k^{-1} \leq 2^k \log L_0^{-1}.$$

Finally,

$$\begin{aligned} \log L_k^{-1} &\geq -1 - 2 - \dots - 2^{k-1} + 2^k \log L_0^{-1} \\ &\geq -2^k + 2^k \log L_0^{-1} = 2^k (\log L_0^{-1} - 1). \end{aligned}$$

Appendix C

Proof of Proposition 2.3.3

If $(\alpha_k, \beta_k) \in B_m$, where m is a positive integer and $m \neq 1$, then

$$\frac{L_{k+1}}{L_k^{\sqrt{2}}} = \frac{1 - (1 - \alpha_k)^2 + \beta_k^2}{(\alpha_k + \beta_k)^{\sqrt{2}}}.$$

To prove the upper bound of the ratio, it suffices to show that

$$\psi(\alpha_k, \beta_k) = 1 - (1 - \alpha_k)^2 + \beta_k^2 - (\alpha_k + \beta_k)^{\sqrt{2}} \leq 0.$$

The second-order partial derivative of ψ with respect to α_k is non-positive:

$$\frac{\partial^2 \psi}{\partial \alpha_k^2} = -2 - \sqrt{2}(\sqrt{2} - 1)(\alpha_k + \beta_k)^{\sqrt{2}-2} \leq 0.$$

Therefore, the minimum of $\partial\psi/\partial\alpha_k$ is on the lines $\alpha_k + \beta_k = 1$ and $(1 - \alpha_k)^2 + \beta_k^2 = 1$. It is easy to show that $\partial\psi/\partial\alpha_k \geq 0$. In consequence, the maximum of ψ is on the lines $\alpha_k + \beta_k = 1$ and $(1 - \alpha_k)^2 + \beta_k^2 = 1$. If $\alpha_k + \beta_k = 1$, then it is easy to see that $\psi = 0$. If $(1 - \alpha_k)^2 + \beta_k^2 = 1$, then $\psi = 2\beta_k^2 - (\alpha_k + \beta_k)^{\sqrt{2}}$. It is easy to show that the maximum value of ψ lies at the intersection of $\alpha_k + \beta_k = 1$ and $(1 - \alpha_k)^2 + \beta_k^2 = 1$, where $\psi = 0$. Hence, the ratio $L_{k+1}/L_k^{\sqrt{2}}$ is upper bounded by 1.

Appendix D

Proof of Proposition 2.3.4

From Proposition 2.3.3 we have, for $k = 0, 1, \dots, m-2$,

$$L_{k+1} = a_k L_k^{\sqrt{2}}$$

for some $a_k \in (0, 1]$. Then for $k = 1, 2, \dots, m-1$,

$$L_k = a_{k-1} \cdot a_{k-2}^{\sqrt{2}} \dots a_0^{\sqrt{2}^{k-1}} L_0^{\sqrt{2}^k},$$

where $a_i \in (0, 1]$ for each i . Hence,

$$\begin{aligned} \log L_k^{-1} &= -\log a_{k-1} - \sqrt{2} \log a_{k-2} - \dots \\ &\quad - \sqrt{2}^{k-1} \log a_0 - \log L_0^{\sqrt{2}^k}. \end{aligned}$$

Since $\log L_0^{-1} > 0$ and $\log a_i \leq 0$ for each i , we have

$$\log L_k^{-1} \geq \sqrt{2}^k \log L_0^{-1}.$$

Therefore, we have

$$\log P_N^{-1} \geq \sqrt{N} \log L_0^{-1}.$$

Appendix E

Proof of Proposition 2.3.5

Because of symmetry, we only have to prove the case where (α_k, β_k) lies in $R_{\mathcal{U}}$. We consider two cases: $(\alpha_k, \beta_k) \in B_1$ and $(\alpha_k, \beta_k) \in B_2 \cap R_{\mathcal{U}}$.

In the first case,

$$\frac{L_{k+2}}{L_k^2} = \frac{(1 - (1 - \alpha_k)^2)^2 + 1 - (1 - \beta_k^2)^2}{(\alpha_k + \beta_k)^2}.$$

To prove the lower bound of the ratio, it suffices to show that

$$\begin{aligned} L_{k+2} - L_k^2 &= (1 - (1 - \alpha_k)^2)^2 + 1 - (1 - \beta_k^2)^2 - (\alpha_k + \beta_k)^2 \\ &= (1 - \alpha_k - \beta_k)((\beta_k - \alpha_k)^3 + 2\alpha_k\beta_k(\beta_k - \alpha_k) \\ &\quad + (\beta_k - \alpha_k)^2 + 2\alpha_k^2) \geq 0. \end{aligned}$$

We have $1 - \alpha_k - \beta_k > 0$ and $\beta_k \geq \alpha_k$ for all $(\alpha_k, \beta_k) \in B_1$, resulting in the above inequality.

To prove the upper bound of the ratio, it suffices to show that

$$L_{k+2} - 2L_k^2 = \alpha_k^4 - 4\alpha_k^3 + 2\alpha_k^2 - 4\alpha_k\beta_k - \beta_k^4 \leq 0.$$

The partial derivative with respect to β_k is

$$\frac{\partial(L_{k+2} - 2L_k^2)}{\partial\beta_k} = -4\alpha_k - 4\beta_k^3 \leq 0,$$

which is non-positive. Therefore, it suffices to consider its values on the curve $\beta_k = \alpha_k$, on which $L_{k+2} - 2L_k^2$ is clearly non-positive.

Now we consider the second case, namely $(\alpha_k, \beta_k) \in B_2 \cap R_{\mathcal{U}}$, which gives

$$\frac{L_{k+2}}{L_k^2} = \frac{1 - (1 - \alpha_k)^4 + \beta_k^4}{(\alpha_k + \beta_k)^2}.$$

To prove the lower bound of the ratio, it suffices to show that

$$\begin{aligned} L_{k+2} - L_k^2 &= (1 - (1 - \alpha_k)^4) + \beta_k^4 - (\alpha_k + \beta_k)^2 \\ &= (1 - \alpha_k - \beta_k)(\alpha_k^3 - \alpha_k^2\beta_k - 3\alpha_k^2 + \alpha_k\beta_k^2 \\ &\quad + 2\alpha_k\beta_k - \beta_k^3 - \beta_k^2 + 4\alpha_k) \geq 0. \end{aligned}$$

Therefore, it suffices to show that

$$\begin{aligned} \phi(\alpha_k, \beta_k) &= \alpha_k^3 - \alpha_k^2\beta_k - 3\alpha_k^2 + \alpha_k\beta_k^2 \\ &\quad + 2\alpha_k\beta_k - \beta_k^3 - \beta_k^2 + 4\alpha_k \geq 0. \end{aligned}$$

The partial derivative with respect to β_k is

$$\frac{\partial \phi}{\partial \beta_k} = -(\alpha_k - \beta_k)^2 - 2\beta_k^2 + 2(\alpha_k - \beta_k) \leq 0.$$

Thus, it is enough to consider the values on the upper boundaries $\sqrt{1 - \beta_k} + \sqrt{\alpha_k} = 1$ and $\alpha_k + \beta_k = 1$.

If $\alpha_k + \beta_k = 1$, then the inequality is trivial, and if $\sqrt{1 - \beta_k} + \sqrt{\alpha_k} = 1$, then

$$L_{k+2} - L_k^2 = 2\alpha_k^2(1 - 2\sqrt{\alpha_k})(2\alpha_k - 6\sqrt{\alpha_k} + 5)$$

and the inequality holds because $\alpha_k \leq \frac{1}{4}$ in region $B_2 \cap R_{\mathcal{U}}$.

The claimed upper bound for the ratio L_{k+2}/L_k^2 can be written as

$$\begin{aligned} L_{k+2} - 2L_k^2 &= (1 - (1 - \alpha_k)^4) + \beta_k^4 - 2(\alpha_k + \beta_k)^2 \\ &= -\alpha_k^4 + 4\alpha_k^3 - 8\alpha_k^2 + 4\alpha_k \\ &\quad - 4\alpha_k\beta_k + \beta_k^4 - 2\beta_k^2 \leq 0. \end{aligned}$$

The partial derivative with respect to β_k is

$$\frac{\partial(L_{k+2} - 2L_k^2)}{\partial \beta_k} = -4\alpha_k + 4\beta_k^3 - 4\beta_k \leq 0.$$

Again, it is sufficient to consider values on the upper boundary of B_1 . Hence,

$$L_{k+2} - 2L_k^2 = 2\beta_k^2 - 2(\alpha_k + \beta_k)^2 \leq 0.$$

Appendix F

Proof of Proposition 2.3.6

In the case where $(\alpha_k, \beta_k) \in B_2 \cap R_{\mathcal{U}}$, from Proposition 2.3.3, we have $L_{k+1} \leq L_k^{\sqrt{2}}$. Moreover, it is easy to show that $L_{k+2} \leq L_{k+1}$. Thus, we have $L_{k+2} \leq L_k^{\sqrt{2}}$.

In the case where $(\alpha_k, \beta_k) \in B_1$, it suffices to prove that

$$\vartheta(\alpha_k, \beta_k) = (1 - (1 - \alpha_k)^2)^2 + 1 - (1 - \beta_k^2)^2 - (\alpha_k + \beta_k)^{\sqrt{2}} \leq 0.$$

We take second-order partial derivative of ϑ with respect to α_k along the lines $\alpha_k + \beta_k = c$ in this region. It is easy to show that the derivative is non-negative:

$$\frac{\partial^2 \vartheta}{\partial \alpha_k^2} = 12((1 - \alpha_k)^2 - \beta_k^2) \geq 0.$$

Therefore, we conclude that the maximum of ϑ lies on the boundaries of this region. If $\alpha_k + \beta_k = 1$, then we have $\vartheta(\alpha_k, \beta_k) = 0$. If $(1 - \alpha_k)^2 + \beta_k^2 = 1$, then we have $\alpha_{k+1} = \beta_{k+1}$. Moreover, if $\alpha_{k+1} = \beta_{k+1}$, then we can show that $L_{k+2} = L_{k+1}$. Hence, it suffices to show that $L_{k+1}/L_k^{\sqrt{2}} \leq 1$ on the line $(1 - \alpha_k)^2 + \beta_k^2 = 1$, which has been proved in Proposition 2.3.3. If $\beta_k = \alpha_k$, then $L_{k+1} = L_k$ and $(\alpha_{k+1}, \beta_{k+1})$ lies on the lower boundary of $R_{\mathcal{L}}$, on which we have $L_{k+2}/L_{k+1}^{\sqrt{2}} \leq 1$. Thus, we have $L_{k+2}/L_k^{\sqrt{2}} \leq 1$.

Appendix G

Proof of Proposition 2.3.7

From Proposition 2.3.6 we have, for $k = 0, 2, \dots, \log N - 2$,

$$L_{k+2} = a_k L_k^{\sqrt{2}}$$

for some $a_k \in (0, 1]$. Then for $k = 2, 4, \dots, \log N$, we have

$$L_k = a_{(k-2)/2} \cdot a_{(k-4)/2}^{\sqrt{2}} \dots a_0^{\sqrt{2}^{(k-2)/2}} L_0^{\sqrt{2}^{k/2}},$$

where $a_i \in (0, 1]$ for each i . Therefore,

$$\begin{aligned} \log L_k^{-1} &= -\log a_{(k-2)/2} - \sqrt{2} \log a_{(k-4)/2} - \dots \\ &\quad - \sqrt{2}^{(k-2)/2} \log a_0 - \log L_0^{\sqrt{2}^{k/2}}. \end{aligned}$$

Since $\log L_0^{-1} > 0$ and $\log a_i \leq 0$ for each i , we have

$$\log L_k^{-1} \geq \sqrt{2}^{k/2} \log L_0^{-1}.$$

Therefore, we have

$$\log P_N^{-1} \geq \sqrt[4]{N} \log L_0^{-1}.$$

Appendix H

Proof of Proposition 2.3.8

The first inequality is equivalent to

$$\begin{aligned} L_{k+1} - L_k^2 &= 1 - (1 - \alpha_k)^2 + \beta_k^2 - (\alpha_k + \beta_k)^2 \\ &= 2\alpha_k(1 - (\alpha_k + \beta_k)) \geq 0, \end{aligned}$$

which holds for all $(\alpha_k, \beta_k) \in \mathcal{U}$.

The second inequality is equivalent to

$$\begin{aligned} L_{k+1} - L_k &= 1 - (1 - \alpha_k)^2 + \beta_k^2 - (\alpha_k + \beta_k) \\ &= (\alpha_k - \beta_k)(1 - (\alpha_k + \beta_k)) \leq 0, \end{aligned}$$

which holds for all $(\alpha_k, \beta_k) \in \mathcal{U}$.

Appendix I

Proof of Theorem 2.3.2

From Proposition 2.3.8, we have

$$L_1 = \tilde{a}L_0^2$$

for some $\tilde{a} \geq 1$. And, by Proposition 2.3.5, the following identity holds.

$$L_{k+2} = a_k L_k^2$$

for $k = 1, 3, \dots, \log N - 2$ and some $a_k \in [1, 2]$. Hence, we can write

$$L_k = \tilde{a}^{2^{(k-1)/2}} \cdot a_{(k-1)/2} \cdot a_{(k-3)/2}^2 \cdots a_1^{2^{(k-3)/2}} L_0^{2^{(k+1)/2}},$$

where $a_i \in [1, 2]$ for each i and $\tilde{a} \geq 1$. Let $k = \log N$, we have

$$\begin{aligned} \log P_N^{-1} &= -2^{(k-1)/2} \log \tilde{a} - \log a_{(k-1)/2} - \dots \\ &\quad - 2^{(k-3)/2} \log a_1 + \sqrt{2N} \log L_0^{-1}. \end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and for each i , $\log a_i \geq 0$. Moreover, $\log \tilde{a} \geq 0$. Hence,

$$\log P_N^{-1} \leq \sqrt{2N} \log L_0^{-1}.$$

It follows by Proposition 2.3.8 that

$$L_k = \tilde{a} L_{k-1}$$

for some $\tilde{a} \in (0, 1]$. By Proposition 2.3.5, we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 2, \dots, \log N - 3$ and some $a_k \in [1, 2]$. Thus,

$$L_k = \tilde{a} \cdot a_{(k-3)/2} \cdot a_{(k-3)/2}^2 \cdots a_0^{2^{(k-3)/2}} L_0^{2^{(k-1)/2}},$$

where $a_i \in [1, 2]$ for each i and $\tilde{a} \in (0, 1]$. Hence,

$$\begin{aligned} \log P_N^{-1} &= -\log \tilde{a} - \log a_{(k-1)/2} - \dots \\ &\quad - 2^{(k-3)/2} \log a_1 + \sqrt{\frac{N}{2}} \log L_0^{-1}. \end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and for each i , $0 \leq \log a_i \leq 1$ and $\log \tilde{a} \leq 0$. Thus,

$$\log P_N^{-1} \geq -\sqrt{\frac{N}{2}} + \sqrt{\frac{N}{2}} \log L_0^{-1} = \sqrt{\frac{N}{2}} (\log L_0^{-1} - 1).$$

Appendix J

Proof of Proposition 2.3.9

The upper bound for L_{k+1}/L_k is trivial. By Proposition 2.2.1, if $(\alpha_{k-1}, \beta_{k-1}) \in B_2 \cap R_{\mathcal{U}}$, then

$$1 \leq \frac{L_k}{L_{k-1}^2} \leq 2;$$

i.e.,

$$\frac{1}{2} \leq \frac{L_{k-1}^2}{L_k} \leq 1,$$

and in consequence of Proposition 2.3.5, if $(\alpha_{k-1}, \beta_{k-1}) \in B_2 \cap R_{\mathcal{U}}$, then

$$1 \leq \frac{L_{k+1}}{L_{k-1}^2} \leq 2.$$

Therefore, we have

$$\frac{1}{2} \leq \frac{L_{k+1}}{L_k}.$$

Appendix K

Proof of Theorem 2.3.3

If $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$ and m is even, then by Proposition 2.3.1, we have

$$L_{m+1} = \tilde{a}L_m^2$$

for some $\tilde{a} \in [1, 2]$.

By Proposition 2.3.5, we have

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 2, \dots, m-2, m+1, \dots, \log N - 2$, and some $a_k \in [1, 2]$. Hence,

$$L_k = a_{(k-1)/2} \cdot a_{(k-3)/2}^2 \dots a_0^{2^{(k-1)/2}} L_0^{2^{(k+1)/2}},$$

where $a_i \in [1, 2]$ for each i .

Let $k = \log N$, we have

$$\begin{aligned} \log P_N^{-1} &= -\log a_{(k-1)/2} - 2 \log a_{(k-3)/2} - \dots \\ &\quad - 2^{(k-1)/2} \log a_0 + \sqrt{2N} \log L_0^{-1}. \end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and for each i , $0 \leq \log a_i \leq 1$. Thus,

$$\log P_N^{-1} \leq \sqrt{2N} \log L_0^{-1}.$$

Finally,

$$\log P_N^{-1} \geq -\sqrt{2N} + \sqrt{2N} \log L_0^{-1} = \sqrt{2N} (\log L_0^{-1} - 1).$$

If $(\alpha_m, \beta_m) \in B_2 \cap R_{\mathcal{U}}$ and m is odd, then by Proposition 2.3.9 we have

$$L_{m+2} = \tilde{a}L_{m+1}$$

for some $\tilde{a} \in [1/2, 1]$.

It follows from Proposition 2.3.5 that

$$L_{k+2} = a_k L_k^2$$

for $k = 0, 2, \dots, m-1, m+2, \dots, \log N - 2$ and some $a_k \in [1, 2]$. Therefore,

$$L_k = a_{(k-3)/2} \cdot a_{(k-3)/2}^2 \dots a_0^{2^{(k-3)/2}} \cdot \tilde{a}^{2^{(k-m-2)/2}} L_0^{2^{(k-1)/2}},$$

where $a_i \in [1, 2]$ for each i and $\tilde{a} \in [1/2, 1]$. Hence,

$$\begin{aligned} \log P_N^{-1} &= -\sqrt{\frac{N}{2^{m+2}}} \log \tilde{a} - \log a_{(k-3)/2} - \dots \\ &\quad - \frac{\sqrt{N/2}}{2} \log a_0 + \sqrt{\frac{N}{2}} \log L_0^{-1}. \end{aligned}$$

Notice that $\log L_0^{-1} > 0$ and for each i , $0 \leq \log a_i \leq 1$ and $-1 \leq \log \tilde{a} \leq 0$. Thus,

$$\log P_N^{-1} \leq \sqrt{\frac{N}{2}} \log L_0^{-1} + \sqrt{\frac{N}{2^{m+2}}}.$$

Finally,

$$\log P_N^{-1} \geq -\sqrt{\frac{N}{2}} + \sqrt{\frac{N}{2}} \log L_0^{-1} = \sqrt{\frac{N}{2}} (\log L_0^{-1} - 1).$$

Appendix L

Proof of Theorem 2.3.4

If $\log N < m$, then this scenario is the same as that of Corollary 2.3.1. Therefore,

$$N (\log L_0^{-1} - 1) \leq \log P_N^{-1} \leq N \log L_0^{-1}.$$

If $\log N \geq m$ and $\log N - m$ is odd, then it takes $(m - 1)$ steps for the system to move into B_1 . After it arrives in B_1 , there is an even number of levels left because $\log N - m$ is odd.

By Proposition 2.3.1, we have

$$L_{k+1} = \tilde{a}_k L_k^2$$

for $k = 0, 1, \dots, m - 2$ and some $\tilde{a}_k \in [1, 2]$, and in consequence of Proposition 2.3.5,

$$L_{k+2} = a_k L_k^2$$

for $k = m - 1, m - 3, \dots, \log N - 2$ and some $a_k \in [1, 2]$. Thus,

$$L_k = a_{(k+m-3)/2} \cdot a_{(k+m-5)/2}^2 \cdots a_0^{2^{(k+m-3)/2}} L_0^{2^{(k+m-1)/2}},$$

where $a_i \in [1, 2]$ for each i .

Let $k = \log N$. Then we obtain

$$\begin{aligned} \log P_N^{-1} &= -\log a_{(k+m-3)/2} - 2 \log a_{(k+m-5)/2} - \cdots \\ &\quad - \frac{\sqrt{2^{m-1}N}}{2} \log a_0 + \sqrt{2^{m-1}N} \log L_0^{-1}. \end{aligned}$$

Note that $\log L_0^{-1} > 0$, and for each i , $0 \leq \log a_i \leq 1$. Thus,

$$\log P_N^{-1} \leq \sqrt{2^{m-1}N} \log L_0^{-1}.$$

Finally,

$$\begin{aligned} \log P_N^{-1} &\geq -\sqrt{2^{m-1}N} + \sqrt{2^{m-1}N} \log L_0^{-1} \\ &= \sqrt{2^{m-1}N} (\log L_0^{-1} - 1). \end{aligned}$$

For the case where $\log N - m$ is even, the proof is similar and it is omitted.

Appendix M

Proof of Proposition 2.3.10

Without loss of generality, we consider the upper half of S , denoted by $S_{\mathcal{U}}$. As we shall see, the image of $S_{\mathcal{U}}$ is exactly the reflection of $S_{\mathcal{U}}$ with respect to the line $\beta = \alpha$ (denoted by $S_{\mathcal{L}}$). We know that $S_{\mathcal{U}} := \{(\alpha, \beta) \in \mathcal{U} | \beta \leq \sqrt{\alpha} \text{ and } \beta \geq 1 - (1 - \alpha)^2\}$.

The image of $S_{\mathcal{U}}$ under f can be calculated by

$$(\alpha', \beta') = f(\alpha, \beta) = (1 - (1 - \alpha)^2, \beta^2),$$

where $(\alpha, \beta) \in \mathcal{U}$. The above relation is equivalent to

$$(\alpha, \beta) = (1 - \sqrt{1 - \alpha'}, \sqrt{\beta'}).$$

Therefore, we can calculate images of boundaries for $R_{\mathcal{U}}$ under f .

The image of the upper boundary $\beta \leq \sqrt{\alpha}$ is

$$\sqrt{\beta'} \leq \sqrt{1 - \sqrt{1 - \alpha'}};$$

i.e.,

$$\alpha' \geq 1 - (1 - \beta')^2,$$

and that of the lower boundary $\beta \geq 1 - (1 - \alpha)^2$ is

$$\sqrt{\beta'} \geq 1 - (1 - (1 - \sqrt{1 - \alpha'}))^2;$$

i.e.,

$$\alpha' \leq \sqrt{\beta'}.$$

The function f is monotone. Hence, images of boundaries of $S_{\mathcal{U}}$ are boundaries of $S_{\mathcal{L}}$. Notice that boundaries of $R_{\mathcal{L}}$ are symmetric with those of $R_{\mathcal{U}}$ about $\beta = \alpha$. We conclude that S is an invariant region.

Bibliography

- [1] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 4, pp. 501–510, Jul. 1981.
- [2] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.
- [3] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.
- [4] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.
- [5] G. Polychronopoulos and J. N. Tsitsiklis, "Explicit solutions for some simple decentralized detection problems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 2, pp. 282–292, Mar. 1990.
- [6] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Asymptotic performance of a censoring sensor network," *IEEE Trans. Inform. Theory*, vol. 53, no. 11, pp. 4191–4209, Nov. 2007.
- [7] P. Willett and D. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 355–361, Mar. 1992.
- [8] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I—Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
- [9] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II—Advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.
- [10] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 2, pp. 469–476, Apr. 1998.
- [11] B. Chen and P. K. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 693–699, Feb. 2005.

- [12] B. Liu and B. Chen, "Channel-optimized quantizers for decentralized detection in sensor networks," *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3349–3358, Jul. 2006.
- [13] B. Chen and P. K. Varshney, "A Bayesian sampling approach to decision fusion using hierarchical models," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1809–1818, Aug. 2002.
- [14] A. Kashyap, "Comments on on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1274–1275, Mar. 2006.
- [15] H. Chen, B. Chen, and P. K. Varshney, "Further Results on the Optimality of the Likelihood-Ratio Test for Local Sensor Decision Rules in the Presence of Nonideal Channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 2, pp. 828–832, Feb. 2009.
- [16] G. Fellouris and G. V. Moustakides, "Decentralized Sequential Hypothesis Testing Using Asynchronous Communication," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 534–548, Jan. 2011.
- [17] J. A. Gubner, L. L. Scharf, and E. K. P. Chong, "Exponential error bounds for binary detection using arbitrary binary sensors and an all-purpose fusion rule in wireless sensor networks," in *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Process.*, Taipei, Taiwan, Apr. 19–24 2009, pp. 2781–2784.
- [18] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part I—Tandem structures," *IEEE Trans. Syst., Man and Cybern.*, vol. 21, no. 5, pp. 1044–1059, Sept./Oct. 1991.
- [19] R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, Jul. 1988.
- [20] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection error probabilities in long tandems," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4767–4771, Oct. 2008.
- [21] J. D. Papastravrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 639–653, Jul. 1992.
- [22] V. V. Veeravalli, "Topics in Decentralized Detection," Ph.D. thesis, University of Illinois at Urbana-Champaign, 1992.
- [23] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: Part II—Tree structures," *IEEE Trans. Syst., Man and Cybern.*, vol. 23, no. 1, pp. 211–221, Jan./Feb. 1993.
- [24] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Data fusion trees for detection: Does architecture matter?," *IEEE Trans. Inform. Theory*, vol. 54, no. 9, pp. 4155–4168, Sept. 2008.
- [25] A. R. Reibman and L. W. Nolte, "Design and performance comparison of distributed detection networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 6, pp. 789–797, Nov. 1987.

- [26] W. P. Tay and J. N. Tsitsiklis, "Error Exponents for Decentralized Detection in Tree Networks," in *Networked Sensing Information and Control*, V. Saligrama, Ed., Springer-Verlag, New York, NY, 2008, pp 73–92.
- [27] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Bayesian detection in bounded height tree networks," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4042–4051, Oct. 2009.
- [28] A. Pete, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks with multiple event structures," *IEEE Trans. Autom. Control*, vol. 39, no. 8, pp. 1702–1707, Aug. 1994.
- [29] O. P. Kreidl and A. S. Willsky, "An efficient message-passing algorithm for optimizing decentralized detection networks," *IEEE Trans. Autom. Control*, vol. 55, no. 3, pp. 563–578, Mar. 2010.
- [30] S. Alhakeem and P. K. Varshney, "A unified approach to the design of decentralized detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 9–20, Jan. 1995.
- [31] Y. Lin, B. Chen, and P. K. Varshney, "Decision fusion rules in multi-hop wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, pp. 475–488, Apr. 2005.
- [32] J. A. Gubner, E. K. P. Chong, and L. L. Scharf, "Aggregation and compression of distributed binary decisions in a wireless sensor network," in *Proc. Joint 48th IEEE Conf. on Decision and Control and 28th Chinese Control Conf.*, Shanghai, P. R. China, Dec. 16-18 2009, pp. 909–913.
- [33] Z. Zhang, A. Pezeshki, W. Moran, S. D. Howard, and E. K. P. Chong, "Error probability bounds for balanced binary fusion trees," *IEEE Trans. Inform. Theory*, to appear. Available from [arXiv:1105.1187v1].
- [34] Z. Zhang, A. Pezeshki, W. Moran, S. D. Howard, and E. K. P. Chong, "Error probability bounds for binary relay trees with crummy sensors," in *Proc. the 2011 Workshop on Defense Applications of Signal Processing (DASP'11)*, The Hyatt Coolum Resort, Coolum, Queensland, Australia, Jul. 10-14, 2011 (Invited Paper).
- [35] Z. Zhang, A. Pezeshki, W. Moran, S. D. Howard, and E. K. P. Chong, "Error probability bounds for binary relay trees with unreliable communication links," in *Proc. of the Asilomar Conference on Signals, Systems, and Computers*, Asilomar Hotel and Conference Grounds, Pacific Grove, California, November 6-9, 2011, to appear.
- [36] Y. Kanoria and A. Montanari, "Subexponential convergence for information aggregation on regular trees," in *Proc. Joint 50th IEEE Conf. on Decision and Control and European Control Conf.*, Orlando, Florida, Dec. 12-15 2011, to appear.
- [37] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Error probability bounds for M -ary relay trees," in *Proc. of International Symposium of Information Theory*, 2012, submitted.
- [38] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Detection Performance of M -ary Relay Trees with Non-binary Message Alphabets," in *Proc. of Statistical Signal Process. Workshop*, 2012, submitted.

- [39] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Submodularity and Optimality of Fusion Rules in Balanced Binary Relay Trees," in *Proc. of 51th IEEE Conf. on Decision and Control*, 2012, submitted.
- [40] P. K. Varshney, *Distributed Detection and Data Fusion*, Springer-Verlag, New York, NY, 1997.
- [41] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley and Sons, New York, NY, 1968.
- [42] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of the approximations for maximizing submodular set functions," in *Math. Programming*, vol. 14, no. 1, pp. 265–294, 1978.
- [43] F. Gianfelici, V. Battistelli, Methods of information geometry (S. Amari, H. Nagaoka, American Mathematical Society and Oxford University Press, 2000) (book review), *IEEE Transactions on Information Theory* 55 (1) (2009) 2905–2906.
- [44] S. Amari, Information geometry of statistical inference - an overview, in: *IEEE Information Theory Workshop*, Bangalore, India, 2002, pp. 86–89.
- [45] S. Amari, H. Nagaoka, *Methods of Information Geometry*, American Mathematical Society and Oxford University Press, New York, 2000.
- [46] C. R. Rao, Information and accuracy attainable in the estimation of statistical parameters, *Bulletin of the Calcutta Mathematical Society* 37 (1945) 81–91.
- [47] N. N. Chentsov, *Statistical Decision Rules and Optimal Inference*, American Mathematical Society, Providence, Rhode Island, 1982.
- [48] B. Efron, Defining the curvature of a statistical problem (with applications to second order efficiency, *The Annals of Statistics* 3 (6) (1975) 1189–1242.
- [49] S. Amari, *Differential-Geometrical Methods of Statistics (Lecture Notes in Statistics)*, Springer, Berlin, Germany, 1985.
- [50] H. Nagaoka, S. Amari, *Differential geometry of smooth families of probability distributions*, METR, University of Tokyo, Japan, 1982.
- [51] R. E. Kass, P. W. Vos, *Geometrical Foundations of Asymptotic Inference*, Wiley, New York, 1997.
- [52] S. Amari, M. Kawanabe, Information geometry of estimating functions in semi parametric statistical models, *Bernoulli* 3 (1997) 29–54.
- [53] S. Amari, K. Kurata, H. Nagaoka, Information geometry of Boltzmann machines, *IEEE Transactions on Neural Networks* 3 (2) (1992) 260–271.
- [54] S. Amari, Information geometry of the EM and em algorithms for neural networks, *Neural Networks* 8 (9) (1995) 1379–1408.
- [55] S. Amari, Natural gradient works efficiently in learning, *Neural Computation* 10 (2) (1998) 251–276.
- [56] S. Amari, Fisher information under restriction of Shannon information, *Annals of the Institute of Statistical Mathematics* 41 (4) (1989) 623–648.

- [57] L. L. Campbell, The relation between information theory and the differential geometry approach to statistics, *Information Sciences* 35 (3) (1985) 199–210.
- [58] S. Amari, Differential geometry of a parametric family of invertible linear systems-Riemannian metric, dual affine connections and divergence, *Mathematical Systems Theory* 20 (1) (1987) 53–82.
- [59] A. Ohara, N. Suda, S. Amari, Dualistic differential geometry of positive definite matrices and its applications to related problems, *Linear Algebra and Its Applications* 247 (1996) 31–53.
- [60] A. Ohara, Information geometric analysis of an interior point method for semidefinite programming, *Geometry in Present Day Science* (1999) 49–74.
- [61] S. Amari, S. Ikeda, H. Shimokawa, Information geometry of α -projection in mean-field approximation, in *Recent Developments of Mean Field Approximation*, M. Oppor, D. Saad, Eds., MIT Press, Cambridge 2000.
- [62] T. Tanaka, Information geometry of mean field approximation, *Neural Computation* 12 (2) (2000) 1951–1968.
- [63] S. Amari, T. S. Han, Statistical inference under multiterminal rate restrictions: a differential geometric approach, *IEEE Transactions on Information Theory* 35 (2) (1989) 217–227.
- [64] S. Amari, Information geometry on hierarchy of probability distributions, *IEEE Transactions on Information Theory* 47 (5) (2001) 1701–1711.
- [65] S. T. Smith, Covariance, subspace, and intrinsic Cramer-Rao bounds, *IEEE Transactions on Signal Processing* 53 (5) (2005) 1610–1630.
- [66] Y. Cheng, X. Wang, B. Moran, Sensor network performance evaluation in statistical manifolds, in: *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, Scotland, 2010.
- [67] X. Wang, Y. Cheng, B. Moran, Bearings-only tracking analysis via information geometry, in: *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, Scotland, 2010.
- [68] X. Wang, B. Moran, Multitarget tracking using virtual measurement of binary sensor networks, in: *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, 2006.
- [69] B. Efron, The geometry of exponential families, *The Annals of Statistics* 6 (2) (1978) 362–376.
- [70] Y. Bar-Shalom, T. E. Fortmann, *Tracking and Data Association*, Academic Press, New York, 1988.
- [71] N. H. Abdel-All, H. N. Abd-Ellah, H. M. Moustafa, Information geometry and statistical manifold, *Chaos, Solitons and Fractals* 15 (1) (2003) 161–172.
- [72] F. Barbaresco, Innovative tools for radar signal processing based on Cartan’s geometry of SPD matrices & information geometry, in: *2008 IEEE Radar Conference*, Rome, Italy, 2008.

- [73] M. L. Menendez, D. Morales, L. Pardo, M. Salicrij, Statistical tests based on geodesic distances, *Applied Mathematics Letters* 8 (1) (1995) 65–69.
- [74] K. M. Carter, R. Raich, W. G. Finn, A. O. Hero, FINE: Fisher information non-parametric embedding, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (11) (2009) 2093–2098.
- [75] V. Arnold, Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l’hydrodynamique des fluides parfaits. (French), *Annales De L’Institut Fourier (Grenoble)* 16 (1) (1966) 319–361.
- [76] P. Petersen, *Riemannian geometry*, Springer, New York, 1998.
- [77] Z. Yang, J. Laaksonen, Principal whitened gradient for information geometry, *Neural Networks* 21 (2-3) (2008) 232–240.
- [78] S. Kobayashi, K. Nomizu, *Foundations of differential geometry*, Wiley-Interscience, New York, 1996.
- [79] L. Astola, L. Florack, Sticky vector fields, and other geometric measures on diffusion tensor images, in: *Proceedings of 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, USA, 2008.
- [80] Y. Ollivier, Ricci curvature of Markov chains on metric spaces, *Journal of Functional Analysis*, 256 (2009) 810–864.
- [81] A. Gray, The volume of a small geodesic ball of a Riemannian manifold, *Michigan Mathematical Journal* 20 (4) (1974) 329–344.
- [82] X. Pennec, Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements, *Journal of Mathematical Imaging and Vision* 25 (2006) 127–154.
- [83] L. Pronzato, E. Walter, Robust experiment design via stochastic approximation, *Mathematical Biosciences* 75 (1) (1985) 103–120.
- [84] S. Amari, Differential geometry of curved exponential families-curvatures and information loss, *The Annals of Statistics* 10 (2) (1982) 357–385.
- [85] A. P. Dawid, Discussions to Efron’s paper, *The Annals of Statistics* 3 (6) (1975) 1231–1234.
- [86] L. T. Madsen, The geometry of statistical model—a generalization of curvature, Technical Report, Danish Medical Research Council, 1979.
- [87] C. T. J. Dodson, H. Matsuzoe, An affine embedding of the gamma manifold, *Applied Sciences* 5 (1) (2003) 7–12.
- [88] B. O. Koopman, On distribution admitting a sufficient statistic, *Transactions of the American Mathematical Society* 39 (3) (1936) 399–409.
- [89] K. Arwini, L. D. Riego, C. T. J. Dodson, Universal connection and curvature for statistical manifold geometry, *Houston Journal of Mathematics* 33 (2007) 145–161.

- [90] K. Arwini, C. T. J. Dodson, Information Geometry: Near Randomness and Near Independence, Springer, Berlin, Germany, 2008.
- [91] M. J. Wainwright, M. I. Jordan, Graphical Models, Exponential Families, and Variational Inference, Now Publishers Inc, Hanover, USA, 2008.
- [92] F. Nielsen, R. Nock, The entropic centers of multivariate normal distributions, in: Proceedings of European Workshop on Computational Geometry (EuroCG), France, 2008, pp. 221–224.